DOCUMENT RESUME

ED 275 706 TM 860 616

AUTHOR Jannarone, Robert J.
TITLE Models for Reflecting Effective Learning Abilities.
INSTITUTION South Carolina Univ., Columbia. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO ONR-86-2
PUB DATE Aug 86
CONTRACT N00014-86-K00817
NOTE 71p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Aptitude; Bayesian Statistics; *Cognitive Ability; *Estimation (Mathematics); Latent Trait Theory; *Learning; Learning Theories; *Mathematical Models; Maximum Likelihood Statistics; Multiple Choice Tests; *Test Theory
IDENTIFIERS Binary Response; Markov Processes; *Rasch Markov Model; Rasch Model

ABSTRACT

A variety of locally dependent models are introduced having individual difference parameters that may be interpreted as reflecting effective learning abilities. One version is a univariate extension of the Rasch model with a Markov property: the probability that a given individual will pass an item depends on previous items only through the immediately preceding item score. Another version is a bivariate extension of the above Rasch Markov model, yielding separable "aptitude" and "learning ability" parameters. For both models, necessary derivatives are presented that are associated with maximum likelihood as well as certain ad hoc estimation procedures. A small number of simulated as well as analytical accuracy results for the estimation procedures are also provided. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275706

TM

RESEARCH REPORT ONR-86-2

MODELS FOR REFLECTING EFFECTIVE LEARNING ABILITIES

Robert J. Jannarone

Psychology Department
University of South Carolina
Columbia, SC 29208

August, 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for any purpose of the United States Government.

This research was sponsored by Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-86-K00817, Authority Identification Number, NR 4421-544.

TM 860 616

BEST COPY AVAILABLE

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): ONR 86-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of South Carolina; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code): Columbia, South Carolina 29208	7b. ADDRESS (City, State, and ZIP Code): 800 N. Quincy St., Code 442 Arlington, VA 22217
8a. NAME OF FUNDING / SPONSORING ORGANIZATION: Personnel and Training Res.; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-86-K00817
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N 42	RR04204	RR0420401	4421-544

11. TITLE (Include Security Classification)

Models for Reflecting Effective Learning Abilities

12. PERSONAL AUTHOR(S) Robert J. Jannarone

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Report	FROM 3/1/86 TO 8/15/86	8/15/86	62

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A variety of locally dependent models are introduced having individual difference parameters that may be interpreted as reflecting effective learning abilities. One version is a univariate extension of the Rasch model with a Markov property: the probability that a given individual will pass an item depends on previous items only through the immediately preceding item score. Another version is a bivariate extension of the above Rasch Markov model, yielding separable "aptitude" and "learning ability" parameters. For both models, necessary derivatives are presented that are associated with maximum likelihood as well as certain ad hoc estimation procedures. A small number of simulated as well as analytical accuracy results for the estimation procedures are also provided.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Charles E. Davis	22b. TELEPHONE (Include Area Code): (202)696-4046	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

Acknowledgements

I wish to thank Sheng-hui Chu for providing intelligent and energetic programming support for this research. Thanks also to Karla Nevils for braving the tedious technical text editing iterations that were necessary for this report.

Introduction

Psychological test theory is often applied when individuals' effective learning abilities, that is their abilities to first learn and then successfully apply new information, are of primary interest. For example, school applicants' chances of succeeding--both in training programs and in subsequent pursuits--are directly related to learning abilities. In such applications, it would be useful to directly measure corresponding individual differences in learning abilities, rather than gleaning them indirectly from aptitude and/or comprehension test scores, as is common practice. More importantly, the availability of individual differences measures of learning abilities could strongly benefit learning and development theorists. Currently, researchers who are interested in learning abilities are instead obliged to use change scores from aptitude measures (Campione, Brown, Ferra, Jones, & Steinberg, 1985), probabilistic models that exclude individual differences in learning (Groner & Spada, 1977), or some other substitute.

Effective learning abilities could be measured directly if it were possible to construct and score items such as the following. Consider a test in which each item included a description of a new concept, the meaning of which could be deduced from the description with some difficulty. As with many standard comprehension test items, the description would be followed by a question regarding the new concept. However, this test would be different in that the question for each item would require not only knowledge of the concept described by the item, but knowledge of concepts introduced in previous items as well. For example, the meaning of some obscure or nonexistent word could be introduced in one item and required, along with other newly introduced information, for passing a subsequent item. Given the

availability of such a test, along with a basis for efficient estimation, the evaluation of effective learning abilities would seem feasible.

Current estimation methods based on locally independent Rasch, logistic, binomial, and classical test models (Lord & Novick, 1968) would be inappropriate for tests of the above type for two reasons. First, such tests would directly violate local independence in that an individual's performance on an item could affect that individual's probability of passing subsequent items. Second, locally independent models would lead to inappropriate latent trait estimates, because they are based on adding individuals' item scores. To explain this second point intuitively, the most direct way to evaluate if a new concept was both learned in one (binary) item and successfully used to answer a subsequent item would be to evaluate the product of the two item scores. From this viewpoint, locally independent models could not directly reflect effective learning abilities, because they involve only (possibly weighted) sums of item scores in calculating latent trait sufficient statistics.

There is a more precise reason why locally independent models cannot reflect effective learning abilities. Jannarone (1986) has introduced a broad family of so-called conjunctive item response models such that each distinct model leads to a distinct sufficient statistic for latent traits. The locally independent models listed above are conjunctive family members having additive sufficient statistics for latent traits. The conjunctive class also includes a variety of models having item product scores involved in their latent trait sufficient statistics. Interestingly, for any model within this family a sufficient statistic for latent traits is additive in the item scores if and only if the model is locally independent. Thus, whenever it would be necessary to compute item product scores in order to

evaluate a construct of interest such as effective learning ability, it would also be necessary to restrict focus to latent trait models and estimation procedures that are locally dependent. (Some recently developed locally independent models for test design that can reflect uniform learning effects in a sample are receiving increased attention (Embretson, 1985a; Fisher and Formann, 1982; Spada & McGaw, 1985)--these models do not estimate distinct individuals' effective learning abilities, however.)

Several models that are locally dependent, hence potentially useful for reflecting effective learning abilities in a single test, have been introduced (Andrich, 1985; Spray and Ackerman, 1986; Jannarone, 1986; Kempf, 1977). However, each among these has its drawbacks. Although estimation procedures for Andrich's (1985) and Kempf's (1977) models have been worked out, these procedures do not estimate distinct individuals' effective learning ability parameters, but instead a single parameter that can only reflect uniform effective learning in a sample. Among the two remaining models, an estimation procedure has been worked out for only one (Jannarone, 1986), but the procedure is both theoretically and practically unsatisfactory.

Purpose. The main purpose of this article is to present a family of viable, locally dependent models for reflecting effective learning abilities, by establishing interpretable parameter/sufficient statistic relationships for the models, along with an efficient estimation framework. An auxiliary purpose is to further the development of workable methods for general conjunctive binary response models, following Jannarone (1986).

Overview. In this section, the key concepts of the article will be introduced in a learning ability assessment setting and with minimal attention to technical details. Most of the concepts that are introduced here will be covered more abstractly and formally in later sections.

In most of what follows, focus will be restricted to the following special case of the hypothetical effective learning ability test introduced earlier. In item 1 a concept would be introduced and tested by a binary question; in item 2 a new concept would be introduced, and the item 2 question would be based on the concept for both items 1 and 2; in item 3 another new concept would be introduced, but the question for item 3 would be based only on the concept for item 3 and the concept for the immediately preceding item; and so on, so that the last question would be based only on the last two concepts introduced in the test. For reasons that will be clarified later, tests having this restricted format will be called Markov chain tests.

For simplicity, suppose first that a single global ability (θ) were related to both aptitude or comprehension performance and effective learning performance on an M-item test. Figure 1 illustrates one such possibility, corresponding to the Markov chain test. As Figure 1 shows, θ could effect individual item scores in two ways: directly, insofar as the concept introduced for an item would be tested by the item; and jointly, insofar as the question for an item would require the concept for a preceding item as well. The direct effect of θ on test performance would correspond to the usual aptitude/comprehension constructs that are associated with tests based on locally independent models. The joint effect, by contrast, would correspond to item linkages reflecting effective learning along with local dependence.

Figure 2 illustrates a second possibility for a Markov chain item response model. For the Figure 2 case, each individual has two latent abilities, γ and δ. As Figure 2 implies, γ values would influence an individual's singular item performance, whereas δ values would effect an individual's conjoint performance on adjacent items.

An analysis of individuals' ability sufficient statistics--given certain effective learning models to be formulated in later sections--may be useful at this point for introducing as well as contrasting the models in Figures 1 and 2 with locally independent models, in terms of observables. Among several item response models that will be introduced for the structures in Figures 1 and 2, two--termed Rasch Markov--models provide simple connections between sufficient statistics for individuals' abilities and their corresponding ability parameters. For the Figure 1 case, a univariate Rasch Markov model will be introduced such that sufficient statistics for θ take the form,

$$t = \sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} x_n x_{n+1} , \qquad (1)$$

where the x_m are binary item scores. For the Figure 2 case, a bivariate Rasch Markov model will be introduced such that sufficient statistics for γ are the usual number-correct scores,

$$g = \sum_{m=1}^{M} x_m , \qquad (2)$$

whereas sufficient statistics for δ take the form,

$$d = \sum_{n=1}^{M-1} x_n x_{n+1} . \qquad (3)$$

In either the Figure 1 or the Figure 2 case, latent traits can be viewed as influencing test scores in two ways for any pair of adjacent items: directly such that missing one item can be neutralized in the additive statistic (2) by passing the other item; and conjointly such that the cross-product statistic (3) records only the passing of both items. Following current terminology, (Embretson, 1985b; Jannarone, 1986; Sympson, 1977), the direct influence of items caused by γ and reflected by (2) will be called

compensatory in the sequel, whereas the conjoint influence caused by δ and reflected by (3) will be called conjunctive.

Table 1 gives the possible contingencies among g and d for a 15-item test, with rows corresponding to the M+1 possible g values and columns corresponding to the M possible d values. (A derivation of the Table 1 contingencies is given in Appendix 1.) As the rows of Table 1 indicate, considerable new information about individuals could be obtained by utilizing either t or (g,d) rather than g alone (given an appropriate measurement model), especially in the middle score ranges. For example, if individuals had a number-correct score of 7, they could be broken down into 6 subgroups if d were used in addition to g, with the potential for increased resolution among the individuals (and corresponding reduced uncertainty with respect to their ability estimates.)

Table 1 also illustrates certain limitations associated with the use of t or (g,d), relative to the use of g alone, namely that g and d tend to be highly correlated, hence highly redundant. In particular, d can never be larger than g-1 (unless g is 0), and d is highly constrained when g takes on values near 0 or M. The tendency for high g,d correlations is so strong that if each of the 2^{15} patterns for a 15-item test were equally likely, then g and d would have a product-moment correlation of about 0.9. The redundancy problem, which implies that little additional information is available from t or (g,d), relative to g alone, is potentially even more serious in settings where γ and δ are highly correlated. In sum, from an information availability viewpoint Table 1 shows that models based on t or on (g,d) have both potential assets and limitations. A method for quantifying these assets and limitations will be outlined later in this section.

A key feature of the models to be presented is that they are locally

dependent. Specifically, for an M-vector of binary item random variables, $\underset{\sim}{X}$, the univariate effective learning models to be introduced do not satisfy

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}|\theta\} = \prod_{m=1}^{M} \Pr\{X_m=x_m|\theta\} , \tag{4}$$

for all θ, whereas the bivariate effective learning models do not satisfy

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}|\gamma,\delta\} = \prod_{m=1}^{M} \Pr\{X_m=x_m|\gamma,\delta\} , \tag{5}$$

for all γ and δ. Figure 3 illustrates conditional and unconditional item response functions for one item based on a particular effective learning model. Because items 4 and 5 in Figure 3 are locally dependent, the conditional item response functions for item 5 are different when responses on item 4 differ. Specifically, the response function for item 5 in Figure 3 corresponds to what will be called a uniformly negative learning transfer case, in that passing item 4 would increase the likelihood of passing item 5.

In practice, it seems likely that uniformly positive learning transfer would occur more often than negative transfer in effective learning tests, in that passing an item would spur an individual toward passing the next item. It seems even more likely that individual differences in learning transfer would also be evident, probably with high ability individuals tending to have more positive learning transfer than low ability individuals. Both the univariate and bivariate effective learning models to be presented will allow for item pairs exhibiting positive, negative, or zero learning transfer, depending on the values of conjoint item parameters for each adjacent item pair. Moreover, the bivariate versions will allow for individual differences in learning transfer as well.

Some Rasch versions of the effective learning models to be introduced also have the Markov property: For univariate models corresponding to Figure 1, it will be shown that

$$\Pr\{X_m=x_m|\theta,x_1,\ldots,x_{m-1}\} = \Pr\{X_m=x_m|\theta,x_{m-1}\},\ m=2,\ldots,M, \qquad (6)$$

whereas for effective learning models corresponding to Figure 2 it will be shown that

$$\Pr\{X_m=x_m|\gamma,\delta,x_1,\ldots,x_{m-1}\} = \Pr\{X_m=x_m|\gamma,\delta,x_{m-1}\},$$

$$m=2,\ldots,M\ . \qquad (7)$$

Some other models will be introduced that satisfy higher-order Markov properties as well.

All of the effective learning models to follow are special cases of the conjunctive item response model family described by Jannarone (1986). One advantage of conjunctive models is that they lead to clear connections between parameters, sufficient statistics, and substantive interpretations, as has already been suggested for effective learning applications. One disadvantage is that efficient estimation procedures for conjunctive models may not exist in general. Efficient procedures are possible for some of the effective learning models to be presented, however, because these models exhibit a certain symmetry. For example, in the Markov models corresponding to Figures 1 and 2, each item will be linked to other items by joint item parameters if and only if the items are adjacent. Because of this symmetric linking of each item (except the first and last) to exactly two other items, it will be possible to derive certain required extensions of the elementary symmetric functions that are used for estimation in the usual Rasch model (Andersen, 1980), thus making efficient estimation feasible.

The necessary first and second derivatives for maximum likelihood estimation will be derived only for the Rasch Markov models to be introduced, which are among the simplest effective learning models possible. Although Rasch Markov models are relatively simple, maximum likelihood estimation

tends to be much slower for them than for the usual Rasch model, because item parameters are dependent. Computational accuracy also appears to be more of a problem. Consequently, an alternative ad hoc estimation procedure that is both fast and accurate will also be presented, having some similarities to the conditional maximum likelihood method for the usual Rasch model (Andersen, 1980).

A major question that must be addressed for these new effective learning models is the extent to which they offer advantages over existing models, if at all. This is especially necessary in light of the earlier-described potential redundancy of effective learning information associated with t and d, relative to the information associated with the usual number-correct score, g. The model comparison approach that will be used compares the validities of Bayes estimators for three competing models. (An explanation for the choice of this comparison approach will be given later.) For the additive Rasch model as well as both univariate and bivariate Rasch Markov models, computer algorithms will be used to evaluate Bayes estimates in the following way. First, test pattern probabilities will be computed, based on the appropriate true item response kernels and discrete latent trait probabilities for the models to be compared. Second, Bayes estimates will be computed analytically, based on each of the prespecified true models and the sufficient statistics associated with each of the assumed models. Finally, Bayes validities, that is population correlations of the Bayes estimates with their corresponding latent traits, will be computed for each assumed model/true model pair, as a basis for comparing the different models.

Table 2 gives the results of the Bayes validity analysis. Focusing on the 15-item results, Table 2 shows that greater validity is possible for tests satisfying a univariate Rasch Markov model (0.93) than for an otherwise

similar additive Rasch model (0.87). However, practically the same validity (0.93) would be obtainable given a Rasch Markov model, if the number-correct statistic for the usual Rasch model were used instead of the optimal statistic, t. This verifies the earlier suggestion that redundancy prevails for d relative to g, hence for t relative to g. Direct comparisons of potential validities for univariate versus bivariate models are more difficult to make, because the bivariate models result in two individual parameters, compared to one individual parameter for univariate models. An ultimate comparison would require knowing how well δ can predict a criterion of interest, over and above γ. However, Table 2 does provide some useful information for bivariate/univariate comparisons. First, assuming the existence of a Rasch model or a univariate Rasch Markov model can lead to reductions in validity if in fact a true bivariate Rasch Markov model holds, especially for low or negative γ, δ correlations (validities of 0.32 and 0.30 vs. 0.71 in the case where $\rho_{\gamma\delta} = -1$). Second, a true additive Rasch Model will yield much more valid estimates of g than an otherwise similar bivariate Rasch Markov model.

In sum, the results of the validity comparison in Table 2 suggest that univariate Rasch Markov tests could be more precise than additive tests of equal length, but the same essential information exists in the number-correct sufficient statistic, g, as in the univariate Rasch Markov statistic, t. The bivariate Rasch Markov results seem more interesting, however, in that if a bivariate model holds and if in fact δ is a valid predictor for effective learning ability, then the bivariate model would offer some advantages over either the additive Rasch model or the univariate Rasch Markov model.

A comprehensive analysis of estimation procedure performance is beyond this article's scope. However, a few results will be given from simulated

test scores in order to verify that the estimation procedures converge and illustrate some of their weaknesses.

Table 3 gives some estimation procedure accuracy figures based on simulated data from 10-item effective learning tests. Test patterns for Table 3 were based on Rasch Markov models having negative transfer. The univariate results correspond to the Figure 1 case, whereas the bivariate results correspond to the Figure 2 case. Although the basis for the Table 3 results requires more explanation later, some of the key results can be pointed out here. For univariate simulations, correlations between estimated and true individual parameter values are relatively high; item parameter estimates are biased toward 0; and individual parameter estimates are slightly biased toward 0. For bivariate simulations, true-estimated individual parameter correlations are relatively low and parameter bias is more evident. Explanations for the estimation performance problems that are suggested by Table 3 as well as possible solutions for future study will be given in later sections.

A detailed description of effective learning models will be given in the remainder of this article. First, the univariate Rasch Markov model will be formulated and its key properties will be derived, followed by a derivation of the necessary formulas for efficient estimation. Second, the closely related bivariate Rasch Markov model and its essential estimation formulas will be presented. Third, the necessary formulas for the previously mentioned validity analysis and estimation results will be given. (A general validity analysis framework will also be provided, because researchers may find it useful to apply the Bayes validity approach in similar studies involving early comparisons among competing models.) Fourth, existing models that are related to Rasch Markov models will be described and several other

effective learning models will be introduced. Finally, some future directions will be suggested.

A Univariate Rasch Markov Model

The Univariate Rasch Markov kernel. The usual Rasch model for item response probabilities may be expressed as,

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\} = \int_{-\infty}^{\infty} \Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\}\,dH(\theta) \qquad (8)$$

where $\underset{\sim}{X}$ is an M-variate binary item response random variable, θ is a latent trait parameter value, and H is the latent trait distribution function. The Rasch Kernel, $\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\}$ may be expressed (Jannarone, 1986) as

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\} = \nu(\theta,\underset{\sim}{\beta})\exp\{\sum_{m=1}^{M}(\theta-\beta_m)x_m\},$$

where the β_m are item difficulty parameters,

$$\nu(\theta,\underset{\sim}{\beta}) = \left[\prod_{m=1}^{M}(1+\exp\{\theta-\beta_m\}\right]^{-1}$$

$$= \left[\sum_{\underset{\sim}{u}\in\chi_M}\exp\{\sum_{m=1}^{M}(\theta-\beta_m)u_m\}\right]^{-1},$$

and

$$\chi_M = \left\{\underset{M\times 1}{\underset{\sim}{u}}:\ u_m = 0,1;\ m=1,\ldots,M\right\}.$$

(The M subscript will be dropped from χ in the sequel.) For the Markov extension to the Rasch kernel, $\underset{\sim}{\beta}$ is extended from an M-vector to a (2M-1)-vector, $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_M,\beta_{12},\beta_{23},\ldots,\beta_{M-1,M})$,

and

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\} =$$

$$\nu(\theta,\underset{\sim}{\beta})\exp\left\{\sum_{m=1}^{M}(\theta-\beta_m)x_m + \sum_{n=1}^{M-1}(\theta-\beta_{n,n+1})x_n x_{n+1}\right\}, \qquad (9)$$

where

$$\nu(\theta,\underset{\sim}{\beta}) = \left[\sum_{\underset{\sim}{u}\in X} \exp\{ \sum_{m=1}^{M} (\theta-\beta_m)u_m + \sum_{n=1}^{M-1} (\theta-\beta_{n,n+1})u_n u_{n+1}\} \right]^{-1}. \qquad (10)$$

Unlike the Rasch kernel, the Rasch Markov kernel does not satisfy the local independence property,

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\} = \prod_{m=1}^{M} \Pr\{X_m=x_m\,|\,\theta,\underset{\sim}{\beta}\},$$

because for any fixed $\underset{\sim}{\beta}$ no $\nu_m(\theta,\beta_m)$ exist satisfying

$$\nu(\theta,\beta) = \prod_{m=1}^{M} \nu_m(\theta,\beta_m),$$

identically in θ and $\underset{\sim}{\beta}$ (Jannarone, 1986).

The Rasch Markov kernel (9) is a member of the so-called conjunctive Rasch kernel family (Jannarone, 1986), from which it follows that if the (2M-1)-vector $\underset{\sim}{\beta}$ is fixed, then

$$t = \sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} x_n x_{n+1} \qquad (11)$$

is a sufficient statistic for θ. Also, (8) may be expressed in the natural multiparameter exponential family (Lehmann, 1983) form,

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\theta,\underset{\sim}{\beta}\} = \nu(\theta,\underset{\sim}{\beta})\exp\left\{\theta\left(\sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} x_n x_{n+1} \right) - \left(\sum_{m=1}^{M} x_m\beta_m + \sum_{n=1}^{M-1} x_n x_{n+1}\beta_{n,n+1} \right) \right\}, \qquad (12)$$

from which it follows that unique maximum likelihood estimates of θ exist for fixed $\underset{\sim}{\beta}$ (except when $t = 0$ or $t = 2M-1$, in which case the likelihood is maximized only as estimates approach $-\infty$ and ∞, respectively).

The log-likelihood function for test scores from a random sample of I individuals is,

$$L = \sum_{i=1}^{I} \log[\Pr\{X_i=x_i \mid \theta_i, \underset{\sim}{\beta}\}]$$

$$= \sum_{i=1}^{I} \{\log[\nu(\theta_i, \underset{\sim}{\beta})] + \sum_{m=1}^{M} (\theta_i-\beta_m)x_{im} + \sum_{n=1}^{M-1} (\theta_i-\beta_{n,n+1})x_{in}x_{i,n+1}\}$$

$$= \sum_{i=1}^{I} \log[\nu(\theta_i, \underset{\sim}{\beta})] + \sum_{i=1}^{I} \theta_i \left(\sum_{m=1}^{M} x_{in}x_{i,n+1} \right) -$$

$$\sum_{m=1}^{M} \beta_m \sum_{i=1}^{I} x_{im} - \sum_{n=1}^{M-1} \beta_{n,n+1} \sum_{i=1}^{I} x_{in}x_{i,n+1}. \qquad (13)$$

It follows again from exponential family theory that if the θ_i are fixed then

$$\left(\sum_{i=1}^{I} x_{i1}, \ldots, \sum_{i=1}^{I} x_{iM}, \sum_{i=1}^{I} x_{i1}x_{i2}, \sum_{i=1}^{I} x_{i2}x_{i3}, \ldots, \sum_{i=1}^{I} x_{i,M-1}x_{iM} \right)$$

are jointly sufficient for $\underset{\sim}{\beta}$ and unique maximum likelihood estimates for $\underset{\sim}{\beta}$ exist (if none of the item sufficient statistics have values of either 0 or I).

For notational and computational simplicity the $\exp\{\theta_i-\beta_m\}$ will sometimes be denoted by λ_{im} in the sequel. Thus, for example, the Rasch Markov kernel (9) may be expressed as

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}_i \mid \underset{\sim}{\Theta}=\theta_i, \underset{\sim}{\beta}\} =$$

$$\nu(\underset{\sim}{\lambda}_i) \left[\prod_{m=1}^{M} \lambda_{im}^{x_{im}} \right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{x_{in}x_{i,n+1}} \right], \qquad (14)$$

where $\underset{\sim}{\lambda}_i$ is a (2M-1)-vector and

$$\nu^{-1}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u} \in \chi} \left[\prod_{m=1}^{M} \lambda_{im}^{u_{im}} \right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_{in}u_{i,n+1}} \right]. \qquad (15)$$

(Individuals' subscripts will be deleted from the λ_{im} in the sequel whenever confusion is unlikely.)

Local dependence and the Markov property. Within the item response context the Markov property would be satisfied if,

$$\Pr\{X_s=x_s \mid X_1=x_1,\ldots,X_{s-1}=x_{s-1};\theta,\underset{\sim}{\beta}\} =$$

$$\Pr\{X_s=x_s \mid X_{s-1}=x_{s-1};\theta,\underset{\sim}{\beta}\},\ s=3,\ldots,M. \qquad (16)$$

We will show that the Markov kernel (9) satisfies (16), first for the case where M = 3 and then for the general case. We will also illustrate local dependence for the case where M = 3.

When M = 3, $\underset{\sim}{\beta}$ as well as $\underset{\sim}{\lambda}$ are 5-vectors and

$$\Pr\{X_1=x_1,X_2=x_2 \mid \theta,\underset{\sim}{\beta}\} = \nu(\underset{\sim}{\lambda}) \sum_{x_3=0}^{1} \lambda_1^{x_1}\lambda_2^{x_2}\lambda_3^{x_3}\lambda_{12}^{x_1x_2}\lambda_{23}^{x_2x_3}$$

$$= \nu(\underset{\sim}{\lambda})\lambda_1^{x_1}\lambda_2^{x_2}\lambda_{12}^{x_1x_2} \sum_{x_3=0}^{1} \left[\lambda_3\lambda_{23}^{x_2}\right]^{x_3},$$

so that

$$\Pr\{X_3=x_3 \mid X_1=x_1,X_2=x_2,\theta,\underset{\sim}{\beta}\} = \frac{\nu(\underset{\sim}{\lambda})\lambda_1^{x_1}\lambda_2^{x_2}\lambda_{12}^{x_1x_2}\left[\lambda_3\lambda_{23}^{x_2}\right]^{x_3}}{\Pr\{X_1=x_1,X_2=x_2 \mid \theta,\underset{\sim}{\beta}\}}$$

$$= \frac{\left[\lambda_3\lambda_{23}^{x_2}\right]^{x_3}}{1+\left[\lambda_3\lambda_{23}^{x_2}\right]^{x_3}}, \qquad (17)$$

which is independent of X_1. Similarly, for s=3,...,M, and m=3,...,

$$\Pr\left\{\bigcap_{m=1}^{s-1}(X_m=x_m \mid \theta,\underset{\sim}{\beta})\right\} = \nu(\underset{\sim}{\lambda}) \sum_{x_s,x_{s+1},\ldots,x_M=0,1} \left[\prod_{m=1}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{n,n+1}^{x_nx_{n+1}}\right]$$

$$= \nu(\underset{\sim}{\lambda})\left[\prod_{m=1}^{s-1}\lambda_m^{x_m}\right]\left[\prod_{n=1}^{s-2}\lambda_{n,n+1}^{x_nx_{n+1}}\right] \times$$

$$\sum_{x_s,\ldots,x_M=0,1}\left[\prod_{m=s}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=s-1}^{M-1}\lambda_{n,n+1}^{x_nx_{n+1}}\right],$$

so that

$$\Pr\left\{X_s=x_s \,\middle|\, \bigcap_{m=1}^{s-1}(X_m=x_m);\theta,\underset{\sim}{\beta}\right\} = \frac{\nu(\underset{\sim}{\lambda})\left[\prod_{m=1}^{s-1}\lambda_m^{x_m}\right]\left[\prod_{n=1}^{s-2}\lambda_{n,n+1}^{x_n x_{n+1}}\right]}{\nu(\underset{\sim}{\lambda})\left[\prod_{m=1}^{s-1}\lambda_m^{x_m}\right]\left[\prod_{n=1}^{s-2}\lambda_{n,n+1}^{x_n x_{n+1}}\right]} \times$$

$$\frac{\sum_{x_{s+1},\ldots,x_M=0,1}\left[\prod_{m=s}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=s-1}^{M-1}\lambda_{n,n+1}^{x_n x_{n+1}}\right]}{\sum_{x_s,\ldots,x_M=0,1}\left[\prod_{m=s}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=s-1}^{M-1}\lambda_{n,n+1}^{x_n x_{n+1}}\right]}, \qquad (18)$$

which after cancellation clearly does not depend on X_1 through X_{s-1}. By symmetry, for s=1,...,M-3,

$$\Pr\{X_s=x_s \,|\, \bigcap_{m=s+1}^{M}(X_m=x_m);\theta,\underset{\sim}{\beta}\}$$

does not depend on x_{s+2} through x_M either.

To illustrate the nature of local dependence when M = 3, it is clear from (17) that X_3 depends on X_2 whenever $\lambda_{23} \neq 1$. In particular, the probability of passing item 3 would be greater if item 2 were passed than if item 2 were failed (positive transfer), whenever $\lambda_{23} > 1$. On the other hand, the probability of passing item 3 would be lover if item 2 were passed than if item 2 were failed (negative transfer), whenever $\lambda_{23} < 1$. In either case, the response function for item 3 could be obtained by evaluating

$$\Pr\{X_3=1\,|\,\theta,\underset{\sim}{\beta}\} = \Pr\{X_2=0\,|\,\theta,\underset{\sim}{\beta}\}\,\frac{\lambda_3}{1+\lambda_3} +$$

$$\Pr\{X_2=1\,|\,\theta,\underset{\sim}{\beta}\}\,\frac{\lambda_3\lambda_{23}}{1+\lambda_3\lambda_{23}}, \qquad (19)$$

provided that the response function for item 1 were known. (Item response function formulas will be developed in the next section.) Figure 3 illustrates conditional and unconditional response functions for a negative

transfer case. (Figure 3 has been discussed earlier.)

More generally, when M=4,5,..., for any given individual X_M depends only on X_{M-1} and X_1 depends only on X_2, whereas X_s depends only on both X_{s-1} and X_{s+1} for s=2,...,M-1. In line with (19), it can be shown that,

$$\Pr\{X_1=1,|\theta,\underset{\sim}{\beta}\} = \Pr\{X_2=0|\theta,\underset{\sim}{\beta}\}\frac{\lambda_1}{1+\lambda_1} +$$

$$\Pr\{X_2=1|\theta,\underset{\sim}{\beta}\}\frac{\lambda_1\lambda_{12}}{1+\lambda_1\lambda_{12}}, \tag{20}$$

$$\Pr\{X_M=1|\theta,\underset{\sim}{\beta}\} = \Pr\{X_{M-1}=0|\theta,\underset{\sim}{\beta}\}\frac{\lambda_M}{1+\lambda_M} +$$

$$\Pr\{X_{M-1}=1|\theta,\underset{\sim}{\beta}\}\frac{\lambda_{M-1,M}\lambda_M}{1+\lambda_{M-1,M}\lambda_M} \tag{21}$$

and for s=2,...,M-1,

$$\Pr\{X_s=1|\theta,\underset{\sim}{\beta}\} = \Pr\{X_{s-1}=0,X_{s+1}=0|\theta,\underset{\sim}{\beta}\}\frac{\lambda_s}{1+\lambda_s} +$$

$$\Pr\{X_{s-1}=0,X_{s+1}=1|\theta,\underset{\sim}{\beta}\}\frac{\lambda_s\lambda_{s+1}}{1+\lambda_s\lambda_{s+1}} +$$

$$\Pr(X_{s-1}=1,X_{s+1}=0|\theta,\underset{\sim}{\beta}\}\frac{\lambda_{s-1}\lambda_s}{1+\lambda_{s-1}\lambda_s} +$$

$$\Pr\{X_{s-1}=1,X_{s+1}=1|\theta,\underset{\sim}{\beta}\}\frac{\lambda_{s-1}\lambda_s\lambda_{s+1}}{1+\lambda_{s-1}\lambda_s\lambda_{s+1}}. \tag{22}$$

It follows from the preceding that the Rasch Markov Model will be locally independent if and only if $\beta_{12} = \beta_{34} = \ldots = \beta_{M-1,M} = \theta$.

Item response functions. In deriving the $\Pr\{X_s=1|\underset{\sim}{\lambda}\}$ (s=1,...,M), we will use the fact that

$$\Pr\{X_s=1|\underset{\sim}{\lambda}\} = 1-\Pr\{X_s=0|\underset{\sim}{\lambda}\}$$

$$= 1-\nu(\underset{\sim}{\lambda}) \sum_{\substack{x_1,\ldots,x_{s-1},\\ x_{s+1},\ldots,x_M=0,1}} \left[\prod_{\substack{m=1\\ m\neq s}}^{M} \lambda_m^{x_m}\right]\left[\prod_{\substack{n=1\\ n\neq s-1,s}}^{M-1} \lambda_{n,n+1}^{x_n x_{n+1}}\right]. \tag{23}$$

For s=1, (23) becomes a

$$\Pr\{X_1=1\,|\,\underset{\sim}{\lambda}\} = 1-\nu(\underset{\sim}{\lambda})/\nu^*(\underset{\sim}{\lambda}^{(1)}), \tag{24}$$

where ν^* corresponds to ν for an (M-1)-item test and

$$\underset{\sim}{\lambda}^{(1)} = (\lambda_2,\ldots,\lambda_M,\lambda_{23},\ldots,\lambda_{M-1,M}).$$

Similarly, the response function for item M is,

$$\Pr\{X_M = 1\,|\,\underset{\sim}{\lambda}\} = 1-\nu(\underset{\sim}{\lambda})/\nu^*(\underset{\sim}{\lambda}^{(M)}), \tag{25}$$

where

$$\underset{\sim}{\lambda}^{(M)} = (\lambda_1,\ldots,\lambda_{M-1},\lambda_{12},\ldots,\lambda_{M-2,M-1}).$$

Also, for s=2,...,M-1,

$$\Pr\{X_s=1\,|\,\underset{\sim}{\lambda}\} = 1-\nu(\underset{\sim}{\lambda})/\nu^*(\underset{\sim}{\lambda}^{(s)}), \tag{26}$$

where

$$\underset{\sim}{\lambda}^{(s)} = (\lambda_1,\ldots,\lambda_{s-1},\lambda_{s+1},\ldots,\lambda_M,\lambda_{12},\ldots,\lambda_{s-2,s-1},1,\lambda_{s+1,s+2},\ldots,\lambda_{M-1,M}).$$

In order to generate Rasch Markov item responses, conditional response functions are required for s=2,...,M, because item responses are locally dependent. From (23), these conditional item response functions are,

$$\Pr\{X_s=x_s\,|\,X_{s-1}=x_{s-1};\theta,\underset{\sim}{\beta}\} =$$

$$\frac{\sum_{x_{s+1},\ldots,x_M=0,1}\left[\lambda_s\lambda_{s-1,s}^{x_{s-1}}\right]^{x_s}\left[\prod_{m=s+1}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=s}^{M-1}\lambda_{n,n+1}^{x_n x_{n+1}}\right]}{\sum_{x_s,\ldots,x_M=0,1}\left[\lambda_s\lambda_{s-1,s}^{x_{s-1}}\right]^{x_s}\left[\prod_{m=s+1}^{M}\lambda_m^{x_m}\right]\left[\prod_{n=s}^{M-1}\lambda_{n,n+1}^{x_n x_{n+1}}\right]}. \qquad (27)$$

But (27) is simply the unconditional item response function for an (M-s+1)-item test having parameters $\lambda_s\lambda_{s-1,s}^{x_{s-1}}, \lambda_{s+1}, \ldots, \lambda_M, \lambda_{s,s+1}, \ldots, \lambda_{M-2,M-1}$, and $\lambda_{M-1,M}$. Thus, (26) can be used to generate the response probabilities for items 2 through M given previous item responses, once the appropriate values are substituted for M and the elements of $\underset{\sim}{\beta}$.

Figure 3 contains conditional and unconditional item response functions that were computed based on the preceding formulas for item 5 of 5-item test having $\beta_1 = \beta_2 = \beta_3 = \beta_{12} = \beta_{23} = \beta_{34} = 0$, $\beta_4 = \beta_5 = -2.5$, and $\beta_{45} = 5.0$. The functions that appear in Figure 3 have been discussed previously.

Maximum likelihood estimation. The formulas to be described in this section can be used for a two-stage Newton-Rapheson estimation scheme of the maximum likelihood type (Lord,1980). With this scheme start values for individual and item parameters are first obtained, after which alternating item and individual parameter estimation steps are performed. During each item (individual) Newton-Rapheson step, the item (individual) parameter estimates are adjusted, while the individual (item) parameters are fixed at their current values. Formulas for the necessary first- and second-order derivatives will be given next.

Beginning with item parameter derivatives, from (13), (14), and (15), we have,

$$\frac{\partial L}{\partial \beta_s} = \sum_{i=1}^{I}\frac{\partial \mathrm{Log}[\nu(\underset{\sim}{\lambda}_i)]}{\partial \beta_s} - \sum_{i=1}^{I} x_{is}$$

$$= \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i) \sum_{\underset{\sim}{u}\in\chi} u_s \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right] - \sum_{i=1}^{I} x_{is}$$

$$= \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)[\nu^{-1}(\underset{\sim}{\lambda}_i) - \nu^{-1}(\lambda_{i1},\ldots,\lambda_{i,s-1},0,\lambda_{i,s+1},\ldots,\lambda_{i,M-1,M})] -$$

$$\sum_{i=1}^{I} x_{is}, \ (s=1,\ldots,M) , \tag{28}$$

with the last step following because

$$\sum_{\underset{\sim}{u}\in\chi} u_s \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in}^{u_n u_{n+1}}\right] = \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=1}} \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in}^{u_n u_{n+1}}\right] ,$$

$$\nu^{-1}(\underset{\sim}{\lambda}_i) = \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=0}} u_s \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right] + \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=1}} \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right],$$

and

$$\nu^{-1}(\lambda_{i1},\ldots,\lambda_{i,s-1},0,\lambda_{i,s+1},\ldots,\lambda_{i,M-1,M})$$

$$= \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=0 \\ \lambda_{is}=0}} \left[\prod_{m=1}^{M} \lambda_{in}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right] \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=1 \\ \lambda_{is}=0}} \left[\prod_{m=1}^{M} \lambda_{in}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right]$$

$$= \sum_{\substack{\underset{\sim}{u}\in\chi \\ u_s=0}} \left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right] , \tag{29}$$

given that $0^1=0$ and $0^0=1$. Denoting (29) by $\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i)$ yields

$$\frac{\partial L}{\partial \beta_s} = I - \sum_{i=1}^{I} x_{is} - \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i). \tag{30}$$

Also, since the $\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i)$ do not depend on $\underset{\sim}{\beta}$,

$$\frac{\partial^2 L}{\partial \beta_s^2} = -\sum_{i=1}^{I} \nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i) \frac{\partial \nu(\underset{\sim}{\lambda}_i)}{\partial \beta_s}$$

$$= -\sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i)[\underset{\sim}{\nu}(\underset{\sim}{\lambda}_i)\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i)-1]. \tag{31}$$

For $t = 1,\ldots,M$, $t \neq s$, $v,w = 1,\ldots,M-1$, $w \neq v-1,v,v+1$, and for $q = 1,\ldots,M, q \neq v, v+1$, similar reasoning leads to

$$\frac{\partial^2 L}{\partial \beta_s \partial \beta_t} = \sum_{i=1}^{I} [\nu^2(\underset{\sim}{\lambda}_i)\nu_{(s)}^{-1}(\underset{\sim}{\lambda}_i)\nu_{(t)}^{-1}(\underset{\sim}{\lambda}_i)-\nu(\underset{\sim}{\lambda}_i)\nu_{(s,t)}^{-1}(\underset{\sim}{\lambda}_i)] \tag{32}$$

$$\frac{\partial L}{\partial \beta_{v,v+1}} = I - \sum_{i=1}^{I} x_{i,v}x_{i,v+1}$$

$$- \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)] \tag{33}$$

$$\frac{\partial^2 L}{\partial \beta_{v,v+1}^2} = \sum_{i=1}^{I} \nu^2(\underset{\sim}{\lambda}_i)[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)] \times$$

$$[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu^{-1}(\underset{\sim}{\lambda}_i)], \tag{34}$$

$$\frac{\partial^2 L}{\partial \beta_{v,v+1}\partial \beta_v} = \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i) \times$$

$$\{\nu(\underset{\sim}{\lambda}_i)[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)]-1\}, \tag{35}$$

$$\frac{\partial^2 L}{\partial \beta_{v,v+1}\partial \beta_{v+1}} = \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i) \times$$

$$\{\nu(\underset{\sim}{\lambda}_i)[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)]-1\}, \tag{36}$$

$$\frac{\partial^2 L}{\partial \beta_{v,v+1}\partial \beta_q} = \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\{\nu(\underset{\sim}{\lambda}_i)\nu_{(q)}^{-1}(\underset{\sim}{\lambda}_i)[\nu_{(v)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(v+1)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)]$$

$$-\nu_{(q,v)}^{-1}(\underset{\sim}{\lambda}_i)-\nu_{(q,v+1)}^{-1}(\underset{\sim}{\lambda}_i)+\nu_{(q,v,v+1)}^{-1}(\underset{\sim}{\lambda}_i)\} \tag{37}$$

$$\frac{\partial^2 L}{\partial\beta_{v,v+1}\partial\beta_{v+1,v+2}} = \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\{\nu^{-1}_{(v,v+1,v+2)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v,v+2)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+1)}(\underset{\sim}{\lambda}_i)+$$

$$\nu(\underset{\sim}{\lambda}_i)[\nu^{-1}_{(v,v+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+1)}(\underset{\sim}{\lambda}_i)] \times$$

$$[\nu^{-1}_{(v+1,v+2)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+2)}(\underset{\sim}{\lambda}_i)]], \qquad (38)$$

and

$$\frac{\partial^2 L}{\partial\beta_{v,v+1}\partial\beta_{w,w+1}} = \sum_{i=1}^{I} \nu(\underset{\sim}{\lambda}_i)\{\nu^{-1}_{(v,v+1,w)}(\underset{\sim}{\lambda}_i)+\nu^{-1}_{(v,v+1,w+1)}(\underset{\sim}{\lambda}_i)+$$

$$\nu^{-1}_{(v,w,w+1)}(\underset{\sim}{\lambda}_i)+$$

$$\nu^{-1}_{(v+1,w,w+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v,w)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v,w+1)}(\underset{\sim}{\lambda}_i)-$$

$$\nu^{-1}_{(v+1,w)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+1,w+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v,v+1,w,w+1)}(\underset{\sim}{\lambda}_i)+$$

$$\nu(\underset{\sim}{\lambda}_i)[\nu^{-1}_{(v,v+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(v+1)}(\underset{\sim}{\lambda}_i)] \times$$

$$[\nu^{-1}_{(w,w+1)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(w)}(\underset{\sim}{\lambda}_i)-\nu^{-1}_{(w+1)}(\underset{\sim}{\lambda}_i)]] . \qquad (39)$$

It can be shown that item parameter derivatives (28) through (39) are well-behaved in that for any given set of item parameter and individual parameter values, (a) each item parameter first-order derivative depends on the data only through its sufficient statistic, (b) each item parameter first-order derivative is a monotone function of the item parameter, (c) none of the second-order derivatives depend on the data, and (d) the matrix of second-order derivatives should be negative definite in a neighborhood of the unique solution point.

Turning next to the individual parameter derivatives when $\underset{\sim}{\beta}$ is fixed, for $i = 1,\ldots,I$ from (13),

$$\frac{\partial L}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\text{Log}[\Pr\{\underset{\sim}{X}_i = \underset{\sim}{x}_i \mid \underset{\sim}{\beta}\}]$$

$$= \frac{\partial}{\partial \theta_i}\left\{\text{Log}[\nu(\theta_i, \underset{\sim}{\beta})] + \theta_i\left(\sum_{m=1}^{M} x_{im} + \sum_{n=1}^{M} x_{in}x_{i,n+1}\right) - \sum_{m=1}^{M} x_{im}\beta_m - \right. \tag{40}$$

$$\left. \sum_{n=1}^{M} x_{in}x_{i,n+1}\beta_{n,n+1}\right\}$$

$$= \frac{1}{\nu(\theta_i, \underset{\sim}{\beta})}\frac{\partial \nu(\theta_i, \underset{\sim}{\beta})}{\partial \theta_i} + \sum_{m=1}^{M} x_{im} + \sum_{n=1}^{M-1} x_{in}x_{i,n+1} ,$$

where

$$\frac{\partial \nu(\theta_i, \underset{\sim}{\beta})}{\partial \theta_i} = \nu^2(\theta_i, \underset{\sim}{\beta}) \sum_{\underset{\sim}{u} \in \chi}\left(\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1}\right)\exp\left\{\sum_{m=1}^{M}(\theta_i - \beta_m)u_m + \right.$$

$$\left. \sum_{n=1}^{M-1}(\theta_i - \beta_{n,n+1})u_n u_{n+1}\right\}$$

$$= -\nu^2(\underset{\sim}{\lambda}_i) \sum_{\underset{\sim}{u} \in \chi}\left[\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1}\right]\left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right]$$

$$= \frac{\partial \nu(\underset{\sim}{\lambda}_i)}{\partial \theta_i} .$$

Thus,

$$\frac{\partial L}{\partial \theta_i} = \sum_{m=1}^{M} x_{im} + \sum_{n=1}^{M-1} x_{in}x_{i,n+1} - \nu(\underset{\sim}{\lambda}_i) \sum_{\underset{\sim}{u} \in \chi}\left[\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1}\right] \times$$

$$\left[\prod_{m=1}^{M} \lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}}\right]$$

$$= \sum_{m=1}^{M} x_{im} + \sum_{n=1}^{M-1} x_{in}x_{i,n+1} - \nu(\underset{\sim}{\lambda}_i) f_1(\underset{\sim}{\lambda}_i) , \tag{41}$$

and

$$\frac{\partial^2 L}{\partial \theta_i^2} = \frac{1}{\nu(\underset{\sim}{\lambda}_i)} \frac{\partial^2 \nu(\underset{\sim}{\lambda}_i)}{\partial \theta_i^2} - \frac{1}{\nu^2(\underset{\sim}{\lambda}_i)} \left(\frac{\partial \nu(\underset{\sim}{\lambda}_i)}{\partial \theta_i} \right)^2$$

$$= \frac{1}{\nu(\underset{\sim}{\lambda}_i)} \frac{\partial^2 \nu(\underset{\sim}{\lambda}_i)}{\partial \theta_i^2} - \nu^2(\underset{\sim}{\lambda}_i) f_1^2(\underset{\sim}{\lambda}_i) ,$$

where

$$\frac{\partial^2 \nu(\underset{\sim}{\lambda}_i)}{\partial \theta_i^2} = \nu^2(\underset{\sim}{\lambda}_i) \sum_{\underset{\sim}{u} \in X} \left[\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1} \right]^2 \times$$

$$\left[\prod_{m=1}^{M} \lambda_{im}^{u_m} \right] \left[\prod_{n=1}^{M-1} \lambda_{in,n+1}^{u_n u_{n+1}} \right] + 2\nu^3(\underset{\sim}{\lambda}_i) f_1^2(\underset{\sim}{\lambda}_i)$$

$$= 2\nu^3(\underset{\sim}{\lambda}_i) f_1^2(\underset{\sim}{\lambda}_i) - \nu^2(\underset{\sim}{\lambda}_i) f_2(\underset{\sim}{\lambda}_i) ,$$

whence

$$\frac{\partial^2 L}{\partial \theta_i^2} = \nu(\underset{\sim}{\lambda}_i) [\nu(\underset{\sim}{\lambda}_i) f_1^2(\underset{\sim}{\lambda}_i) - f_2(\underset{\sim}{\lambda}_i)] . \tag{42}$$

Finally, for $i \neq j$,

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = 0. \tag{43}$$

The computations for (28) through (43) are straightforward with the exception of evaluating the $\nu(\underset{\sim}{\lambda}_i)$, $f_1(\underset{\sim}{\lambda}_i)$ and $f_2(\underset{\sim}{\lambda}_i)$, which would entail adding 2^M terms for each if they were computed directly. Fortunately, ν, f_1, and f_2 can be computed much more efficiently by using certain inductive relationships among their terms, which will be described next. These are related to (but distinct from) the so-called elementary symmetric relationships, which are widely used in Rasch model conditional maximum

likelihood estimation (Anderson, 1982). For any $\underset{\sim}{\lambda}$ having real-valued elements,

$$\nu^{-1}(\underset{\sim}{\lambda}) = \sum_{\underset{\sim}{u}\in\chi} \left(\prod_{m=1}^{M} \lambda_m^{u_m} \right) \left(\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}} \right)$$

$$= 1 + \sum_{i=1}^{M} \sum_{j=1}^{M-i+1} \tau_{ij}, \tag{44}$$

where

$$\tau_{1j} = \lambda_j, \; j=1,\ldots,M, \tag{45}$$

$$\tau_{i1} = \tau_{i-1,1}\lambda_{i-1,i}\lambda_i, \; i=2,\ldots,M, \tag{46}$$

and

$$\tau_{ij} = \left(\sum_{k=1}^{j-1} \tau_{i-1,k}+\tau_{i-1,j}\lambda_{i+j-2,i+j-1} \right)\lambda_{i+j-1},$$

$$i=2,\ldots,M-1, \; j=2,\ldots,M-i+1. \tag{47}$$

Next,

$$f_1(\underset{\sim}{\lambda}) = \sum_{\underset{\sim}{u}\in\chi} \left(\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1} \right) \left(\prod_{n=1}^{M} \lambda_m^{u_m} \right) \left(\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}} \right)$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{M-i+1} \psi_{ij}, \tag{48}$$

where

$$\psi_{1j} = \lambda_j, \; j=1,\ldots,M, \tag{49}$$

$$\psi_{i1} = \psi_{i-1,1}\lambda_i\lambda_{i-1,i}+2\tau_{i,1}, \; i=2,\ldots,M, \tag{50}$$

$$\psi_{ij} = [(2\tau_{i-1,j}+\psi_{i-1,j})\lambda_{i+j-2,i+j-1} +$$

$$\sum_{k=1}^{j-1} (\psi_{i-1,k}+\tau_{i-1,k})]\lambda_{i+j-1}, \; i=2,\ldots,M-1, \; j=2,\ldots,M-i+1. \tag{51}$$

Finally,

$$f_2(\underset{\sim}{\lambda}) = \sum_{\underset{\sim}{u}\in\chi}\left(\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1}\right)^2 \left(\prod_{m=1}^{M} \lambda_m^{u_m}\right)\left(\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}}\right) \cdot$$

$$= f_1(\underset{\sim}{\lambda}) + 2\sum_{i=2}^{M}\sum_{j=1}^{M-i+1} \omega_{ij} , \tag{52}$$

where

$$\omega_{21} = 3\lambda_1\lambda_2\lambda_{12} , \tag{53}$$

$$\omega_{2j} = \lambda_{j+1}\sum_{k=1}^{j-1}\lambda_{k1} + 3\lambda_j\lambda_{j+1}\lambda_{j,j+1} , \quad j=2,\ldots,M-1 , \tag{54}$$

$$\omega_{i1} = (\omega_{i-1,1} + 2\psi_{i-1,1} + \tau_{i-1,1})\lambda_i\lambda_{i-1,i} , \quad i=3,\ldots,M, \tag{55}$$

and

$$\omega_{ij} = \left[\sum_{k=1}^{j-1}(\omega_{i-1,k} + \psi_{i-1,k}) + (\omega_{i-1,j} + 2\psi_{i-1,j} + \tau_{i-1,j})\lambda_{i+j-2,i+j-1}\right] \times$$

$$\lambda_{i+j-1}, \quad i=3,\ldots,M-1, \quad j=2,\ldots,M-i+1. \tag{56}$$

Regarding identifiability, the Rasch Markov kernel (9) is unidentified as it stands in that if the same constant were added to the item parameters as well as the individual parameters then all observable probabilities would be unaffected. As with the additive Rasch model, it is easy to identify (9) simply by fixing one of the parameters during the estimation process.

An ad hoc supplement to maximum likelihood estimation. Maximizing algorithms based on the above partial derivatives tend to be slow, because each iteration involves many symmetric-like function calculations, especially for β maximization. Moreover, certain well-known symmetric function computing innacuracies (Gustaffson, 1980) hold for these models as well. The following ad hoc procedure yields relatively fast "conditional maximum likelihood estimates", $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\theta}}$, that are also accurate. The procedure, which

may be used either to provide start values for--or as an alternative to--direct maximum likelihood, resembles the well-known conditional maximum likelihood method for the Rasch model (Andersen, 1980; Rasch, 1980), in that item parameter estimates are conditioned on a given individual sufficient statistic value. The procedure differs from conditional maximum likelihood, however, in that further conditioning is used in order to remove interitem dependencies.

The first step in the ad hoc procedure is to preselect some individual sufficient statistic value--for example by identifying those I_{MF} individuals in the sample who have the most frequently occurring sufficient statistic value, t_{MF}--and to set θ at zero for each such individual. That subsample is then used to obtain "conditional maximum likelihood" estimates for $\underset{\sim}{\beta}$, after which maximum likelihood estimates for θ are obtained corresponding to each of the 2M-2 remaining statistic values.

The ad hoc $\underset{\sim}{\beta}$ estimates are based on the following formulas. By developments similar to (19) through (22), it can be shown that

$$\Pr\{X_1=x_1 \,|\, X_2=0, \theta=0\} = \frac{\exp\{-\beta_1 x_1\}}{1+\exp\{-\beta_1\}} , \tag{57}$$

$$\Pr\{X_m=x_m \,|\, X_{m-1}=0, X_{m+1}=0, \theta=0\} =$$

$$\frac{\exp\{-\beta_m x_m\}}{1+\exp\{-\beta_m\}} , \quad m=2,\ldots,M-1 , \tag{58}$$

$$\Pr\{X_M=x_M \,|\, X_{M-1}=0, \theta=0\} = \frac{\exp\{-\beta_M x_M\}}{1+\exp\{-\beta_M\}} , \tag{59}$$

$$\Pr\{X_1=x_1, X_2=x_2 \,|\, X_3=0, \theta=0\} =$$

$$\frac{\exp\{-\beta_1 x_1-\beta_2 x_2-\beta_{12} x_1 x_2\}}{1+\exp\{-\beta_1\}+\exp\{-\beta_2\}+\exp\{-\beta_1-\beta_2-\beta_{12}\}} , \tag{60}$$

$$\Pr\{X_m=x_m, X_{m+1}=x_{m+1} \,|\, X_{m-1}=0, X_{m+2}=0, \theta=0\} =$$

$$\frac{\exp\{-\beta_m x_m-\beta_{m+1}x_{m+1}-\beta_{m,m+1}x_m x_{m+1}\}}{1+\exp\{-\beta_m\}+\exp\{-\beta_{m+1}\}+\exp\{-\beta_m-\beta_{m+1}-\beta_{m,m+1}\}}\ ,\ m=2,\ldots,M-1 \quad (61)$$

and

$$\Pr\{X_{M-1}=x_{M-1},X_M=x_M|X_{M-2}=0,\Theta=0\} =$$

$$\frac{\exp\{-\beta_{M-1}x_{M-1}-\beta_M x_M-\beta_{M-1,M}x_{M-1}x_M\}}{1+\exp\{-\beta_{M-1}\}+\exp\{-\beta_M\}+\exp\{-\beta_{M-1}-\beta_M-\beta_{M-1,M}\}}\ . \quad (62)$$

Relatively simple log-likelihood functions can be derived from (57) through (62), for example,

$$L_m\left[\sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+1}=0\\ t_i=t_{MF}}}^{I} x_{im}\,|\beta_m\right] = -\beta_m \sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+1}=0\\ t_i=t_{MF}}}^{I} x_{im} \quad -N[t_i=t_{MF},x_{i,m-1}=x_{i,m+1}=0]\ \times$$

$$\log[1+\exp\{-\beta_m\}],\ m=2,\ldots,M-1\ , \quad (63)$$

and

$$L_{m,m+1}\left[\sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+2}=0\\ t_i=t_{MF}}}^{I} (x_{im}x_{i,m+1})\,|\beta_m\right] = -\beta_m \sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+2}=0\\ t_i=t_{MF}}}^{I} x_{im}$$

$$-\beta_{m+1}\sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+1}=0\\ t_i=t_{MF}}}^{I} x_{i,m+1} \quad -\beta_{m,m+1}\sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+2}=0\\ t_i=t_{MF}}}^{I} x_{im}x_{i,m+1}$$

$$-N[t_i=t_{MF},x_{i,m-1}=x_{i,m+2}=0]\ \log[1+\exp\{-\beta_m\}+\exp\{-\beta_{m+1}\}+$$

$$\exp\{-\beta_m-\beta_{m+1}-\beta_{m,m+1}\}],\ m=2,\ldots,M-2. \quad (64)$$

From (63) we have, for $m=2,\ldots,M-1$,

$$\frac{\partial L_m}{\partial \beta_m} = N[t_i=t_{MF},x_{i,m-1}=x_{i,m+1}=0]\ \frac{\exp\{-\beta_m\}}{1+\exp\{-\beta_m\}} - \sum_{\substack{i=1\\ x_{i,m-1}=x_{i,m+1}=0\\ t_i=t_{MF}}}^{I} x_{im} \quad (65)$$

and

$$\frac{\partial^2 L_m}{\partial \beta_m^2} = -N[t_i=t_{MF}, x_{i,m-1}=x_{i,m+1}=0]/(1+\exp\{-\beta_m\})^2, \tag{66}$$

whereas from (64) we have for m=2,..., M-2,

$$\frac{\partial L_{m,m+1}}{\partial \beta_{m,m+1}} = N[t_i=t_{MF}, x_{i,s-1}=x_{i,s+2}=0] \times \frac{\exp\{-\beta_m-\beta_{m+1}-\beta_{m,m+1}\}}{1+\exp\{-\beta_m\}+\exp\{-\beta_{m+1}\}+\exp\{-\beta_m-\beta_{m+1}-\beta_{m,m+1}\}} - \sum_{\substack{i=1 \\ x_{i,s-1}=x_{i,s+1}=0 \\ t=t_{MF}}}^{I} x_{im}x_{i,m+1}, \tag{67}$$

and

$$\frac{\partial^2 L_{m,m+1}}{\partial \beta_{m,m+1}^2} = -N[t_i=t_{MF}, x_{i,m-1}=x_{i,m+2}=0]/(1+\exp\{-\beta_m\}+\exp\{-\beta_{m+1}\}+\exp\{-\beta_m-\beta_{m+1}-\beta_{m,m+1}\})^2 . \tag{68}$$

For $\underset{\sim}{\beta}$ estimation, the ad hoc approach entails using steepest ascent and/or Newton-Rapheson iterations to find first the $\hat{\beta}_m$ that maximize (63) and then the $\hat{\beta}_{m,m+1}$ that maximize (64), given that $\beta_m = \hat{\beta}_m$ and $\beta_{m+1} = \hat{\beta}_{m+1}$.

A Bivariate Rasch Markov Model

The following model is related to a bivariate model introduced by James Laughlin (Laughlin & Jannarone, 1986; Jannarone & Laughlin, 1986).

The bivariate Rasch Markov kernel. For the bivariate version of the Rasch kernel, $\underset{\sim}{\beta}$ remains the same as the (2M-1)-vector for the univariate version, but θ is replaced $\underset{\sim}{\theta} = (\gamma, \delta)$. The bivariate Rasch Markov kernel is given by,

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\underset{\sim}{\theta},\underset{\sim}{\beta}\} = \nu(\underset{\sim}{\theta},\underset{\sim}{\beta})\exp\left\{\sum_{m=1}^{M}(\gamma-\beta_m)x_m+\sum_{n=1}^{M-1}(\delta-\beta_{n,n+1})x_n x_{n+1}\right\}, \qquad (69)$$

where

$$\nu(\underset{\sim}{\theta},\underset{\sim}{\beta}) = \left[\sum_{\underset{\sim}{u}\in\chi}\exp\left\{\sum_{m=1}^{M}(\gamma-\beta_m)u_n+\sum_{n=1}^{M-1}(\delta-\beta_{n,n+1})u_n u_{n+1}\right\}\right]^{-1}.$$

As with the univariate version, the bivariate kernel (69) may be expressed in natural exponential family form,

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\,|\,\underset{\sim}{\theta},\underset{\sim}{\beta}\}$$

$$= \nu(\underset{\sim}{\theta},\underset{\sim}{\beta})\exp\left\{\gamma\sum_{m=1}^{M}x_m+\delta\sum_{n=1}^{M-1}x_n x_{n+1}-\sum_{m=1}^{M}x_m\beta_m+\sum_{n=1}^{M-1}x_n x_{n+1}\beta_{n+1}\right\}, \qquad (70)$$

from which it follows that (g,d) are jointly sufficient for (γ,δ) given fixed $\underset{\sim}{\beta}$, where

$$g = \sum_{m=1}^{M}x_m$$

and

$$d = \sum_{n=1}^{M-1}x_n x_{n+1}.$$

It also follows that unique maximum likelihood estimates for (γ,δ) exist whenever g and d do not take on boundary values. The preceding developments for the univariate version--from (14) through (39)--are identical for the bivariate case with one exception. For the bivariate case θ_i should be replaced by γ_i in all terms involving the β_m and by δ_i in all terms involving the $\beta_{n,n+1}$. Thus, for the bivariate case the Markov property holds, as well as the previously described univariate $\underset{\sim}{\beta}$ derivatives for maximum likelihood estimation.

For the bivariate case, item response functions are based on γ and δ instead of θ. The univariate item response functions (17) through (19) apply for the bivariate case as well, once the first M elements for $\underset{\sim}{\lambda}$ are based on γ and the last M-1 are based on δ.

For the remainder of this section, the $\underset{\sim}{\lambda}_i$ will be based on γ_i and δ_i instead of θ_i, as in the previous paragraph.

Estimation. The necessary (γ, δ) derivatives for fixed $\underset{\sim}{\beta}$ will be outlined next. For $\underset{\sim}{\theta}_i=(\gamma_i, \delta_i)$ $(i=1,\ldots,I)$, from the bivariate version of (12) we have,

$$\frac{\partial L}{\gamma_i} = \frac{\partial}{\partial \gamma_i} \left\{ \log\left[\nu(\underset{\sim}{\theta}_i, \underset{\sim}{\beta})\right] + \gamma_i \sum_{m=1}^{M} x_{im}\beta_m + \delta_i \sum_{n=1}^{M-1} x_{in}x_{i,n+1} - \sum_{m=1}^{M} x_{im}\beta_m - \sum_{n=1}^{M-1} x_{in}x_{i,n+1}\beta_{n,n+1} \right\}$$

$$= \frac{1}{\nu(\underset{\sim}{\theta}_i, \underset{\sim}{\beta})} \frac{\partial \nu(\underset{\sim}{\theta}_i, \underset{\sim}{\beta})}{\partial \gamma_i} + \sum_{m=1}^{M} x_{im} ,$$

which is quite similar to its univariate counterpart and amenable to nearly the same treatment. A development parallel to that for (40) through (56) yields,

$$\frac{\partial L}{\partial \gamma_i} = \sum_{m=1}^{M} x_{im} - \nu(\underset{\sim}{\lambda}_i) f_{10}(\underset{\sim}{\lambda}_i) , \tag{71}$$

$$\frac{\partial L}{\partial \delta_i} = \sum_{n=1}^{M-1} x_{in}x_{i,n+1} - \nu(\underset{\sim}{\lambda}_i) f_{01}(\underset{\sim}{\lambda}_i) , \tag{72}$$

$$\frac{\partial^2 L}{\partial \gamma_i^2} = \nu(\underset{\sim}{\lambda}_i)[\nu(\underset{\sim}{\lambda}_i) f_{10}^2(\underset{\sim}{\lambda}_i) - f_{20}(\underset{\sim}{\lambda}_i)] , \tag{73}$$

$$\frac{\partial^2 L}{\partial \delta_i^2} = \nu(\underset{\sim}{\lambda}_i)[\nu(\underset{\sim}{\lambda}_i) f_{01}^2(\underset{\sim}{\lambda}_i) - f_{02}(\underset{\sim}{\lambda}_i)] , \tag{74}$$

and

$$\frac{\partial^2 L}{\partial \gamma_i \partial \delta_i} = \nu(\underset{\sim}{\lambda}_i)[f_{11}(\underset{\sim}{\lambda}_i)-\nu(\underset{\sim}{\lambda}_i)f_{10}(\underset{\sim}{\lambda}_i)f_{01}(\underset{\sim}{\lambda}_i)] \ , \tag{75}$$

where

$$f_{10}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u}\in X}\left[\sum_{m=1}^{M} u_m\right]\left[\prod_{m=1}^{M}\lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{in,n+1}^{u_n u_{n+1}}\right] , \tag{76}$$

$$f_{01}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u}\in X}\left[\sum_{n=1}^{M-1} u_n u_{n+1}\right]\left[\prod_{m=1}^{M}\lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{in,n+1}^{u_n u_{n+1}}\right] , \tag{77}$$

$$f_{20}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u}\in X}\left[\sum_{m=1}^{M} u_m\right]^2\left[\prod_{m=1}^{M}\lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{in,n+1}^{u_n u_{n+1}}\right] , \tag{78}$$

$$f_{02}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u}\in X}\left[\sum_{m=1}^{M-1} u_n u_{n+1}\right]^2\left[\prod_{m=1}^{M}\lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{in,n+1}^{u_n u_{n+1}}\right] , \tag{79}$$

and

$$f_{11}(\underset{\sim}{\lambda}_i) = \sum_{\underset{\sim}{u}\in X}\left[\sum_{m=1}^{M} u_m\right]\left[\sum_{n=1}^{M-1} u_n u_{n+1}\right]\left[\prod_{m=1}^{M}\lambda_{im}^{u_m}\right]\left[\prod_{n=1}^{M-1}\lambda_{in,n+1}^{u_n u_{n+1}}\right] . \tag{80}$$

Also, for $i \neq j$,

$$\frac{\partial^2 L}{\partial \gamma_i \partial \gamma_j} = \frac{\partial^2 L}{\partial \delta_i \partial \delta_j} = \frac{\partial^2 L}{\partial \gamma_i \partial \delta_j} = 0. \tag{81}$$

As in the univariate case it is necessary to inductively evaluate certain functions for the bivariate case, namely equations (76) through (80). The required functions for evaluating the $f_{10}(\underset{\sim}{\lambda}_i)$ and $f_{01}(\underset{\sim}{\lambda}_i)$ are,

$$f_{\underset{\sim}{c}1} = \sum_{\underset{\sim}{u}\in X}\left(\sum_{m=1}^{M} c_m u_m + \sum_{n=1}^{M-1} c_{n,n+1} u_n u_{n+1}\right)\left(\prod_{m=1}^{M}\lambda_m^{u_m}\right)\left(\prod_{n=1}^{M-1}\lambda_{n,n+1}^{u_n u_{n+1}}\right)$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{M-i+1} \underset{\sim}{\psi}_{cij}, \tag{82}$$

where

$$\underset{\sim}{\psi}_{c1} = c_j \lambda_j, \quad j=1,\ldots,M, \tag{83}$$

$$\underset{\sim}{\psi}_{ci1} = \underset{\sim}{\psi}_{ci-1,1} \lambda_i \lambda_{i-1,i} + (c_i + c_{i-1,i}) \tau_{i,1}, \quad i=2,\ldots,M, \tag{84}$$

and

$$\underset{\sim}{\psi}_{ci,j} = \left\{ \left[\left(c_{i+j-1} + c_{i+j-2,i+j-1} \right) \tau_{i-1,j} + \underset{\sim}{\psi}_{ci-1,j} \right] \lambda_{i+j-2,i+j-1} + \sum_{k=1}^{j-1} \left(\underset{\sim}{\psi}_{ci-1,k} + c_{i+j-1} \tau_{i-1,k} \right) \right\} \lambda_{i+j-1}, \quad i=2,\ldots,M-1, \; j=2,\ldots,M-i+1. \tag{85}$$

For evaluating the $f_{10}(\underset{\sim}{\lambda}_i)$, the c_m would be set to 1 in (82) and the $c_{n,n+1}$ would be set to 0, whereas the c_m would be set to 0 and the $c_{n,n+1}$ would be set to 1 for evaluating the $f_{01}(\underset{\sim}{\lambda}_i)$. Similarly, the $f_{20}(\underset{\sim}{\lambda}_i)$ and the $f_{02}(\underset{\sim}{\lambda}_i)$ would be evaluated by letting

$$\underset{\sim}{f}_{c2} = \sum_{\underset{\sim}{u} \in \chi} \left[\sum_{m=1}^{M} c_m u_m + \sum_{n=1}^{M-1} c_{n,n+1} u_n u_{n+1} \right]^2 \left[\prod_{m=1}^{M} \lambda_m^{u_m} \right] \left[\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}} \right]$$

$$= f_{(c_1^2,\ldots,c_{M-1,M}^2)1} + 2 \sum_{i=2}^{M} \sum_{j=1}^{M-i+1} \underset{\sim}{\omega}_{cij}, \tag{86}$$

where $f_{(c_1^2,\ldots,c_{M-1,M}^2)1}$ denotes $\underset{\sim}{f}_{c1}$ with the c_M replaced by c_M^2,

$$\underset{\sim}{\omega}_{c21} = (c_1 c_2 + c_1 c_{12} + c_2 c_{12}) \lambda_1 \lambda_2 \lambda_{12}, \tag{87}$$

$$\underset{\sim}{\omega}_{c2j} = c_{j+1} \lambda_{j+1} \sum_{k=1}^{j-1} c_k \lambda_k + (c_j c_{j+1} + c_j c_{j,j+1} + c_{j+1} c_{j,j+1}) \times$$

$$\lambda_j \lambda_{j+1} \lambda_{j,j+1}, \quad j=2,\ldots,M-1 , \tag{88}$$

$$\underset{\sim}{\omega}_{ci1} = \left[\underset{\sim}{\omega}_{ci-1,1}{}^{+(c_i+c_{i-1,i})}\underset{\sim}{\psi}_{ci-1,1}{}^{+c_i c_{i-1,i}\tau_{i-1,1}}\right]\lambda_i \lambda_{i-1,i},$$

$$i=3,\ldots,M , \tag{89}$$

and

$$\underset{\sim}{\omega}_{cij} = \left\{ \sum_{k=1}^{j-1}\left(\underset{\sim}{\omega}_{ci-1,k}{}^{+c_{i+j-1}}\underset{\sim}{\psi}_{ci-1,k}\right) + \right.$$

$$\left[\underset{\sim}{\omega}_{ci-1,j}{}^{+c_{i+j-1}c_{i+j-2,i+j-1}\tau_{i-1,j}{}^{+(c_{i+j-1}+c_{i+j-2,i+j-1})}}\underset{\sim}{\psi}_{ci-1,j}\right] \times$$

$$\left. \lambda_{i+j-2,i+j-1}\right\} \lambda_{i+j-1}, \quad i=3,\ldots M-1, \; j=2,\ldots,M-i+1. \tag{90}$$

Finally, the $f_{11}(\underset{\sim}{\lambda}_i)$ can be evaluated by noting that

$$\sum_{\underset{\sim}{u}\in\chi}\left[\sum_{m=1}^{M} u_m + \sum_{n=1}^{M-1} u_n u_{n+1}\right]^2 \left[\prod_{m=1}^{M} \lambda_m^{u_m}\right]\left[\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}}\right]$$

$$= \sum_{\underset{\sim}{u}\in\chi}\left[\sum_{m=1}^{M} u_m\right]^2 \left[\prod_{m=1}^{M} \lambda_m^{u_m}\right]\left[\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}}\right] +$$

$$2 \sum_{\underset{\sim}{u}\in\chi}\left[\sum_{m=1}^{M} u_m\right]\left[\sum_{n=1}^{M-1} u_n u_{n+1}\right] \times$$

$$\left[\prod_{m=1}^{M} \lambda_m^{u_1}\right]\left[\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}}\right] + \sum_{\underset{\sim}{u}\in\chi}\left[\sum_{n=1}^{M-1} u_n u_{n+1}\right]^2 \times$$

$$\left[\prod_{m=1}^{M} \lambda_m^{u_n}\right]\left[\prod_{n=1}^{M-1} \lambda_{n,n+1}^{u_n u_{n+1}}\right] , \tag{91}$$

whence

$$f_{11}(\underset{\sim}{\lambda}) = [f_{12}(\underset{\sim}{\lambda}) - f_{20}(\underset{\sim}{\lambda}) - f_{02}(\underset{\sim}{\lambda})]/2, \tag{92}$$

where $f_{\underset{\sim}{12}}$ denotes $f_{\underset{\sim}{c2}}$ with the c_m set to 1.

For the ad hoc method, once item parameters are estimated estimates of the individual parameters are obtained with item parameters fixed at their estimated values, as in the maximum likelihood method. However, in conrast with the maximum likelihood approach, no subsequent iterations are performed for the ad hoc method.

The previously described ad hoc procedure for $\underset{\sim}{\beta}$ estimation in the univariate case may be directly applied to the bivariate case as well, except that more than one parameter must be fixed in order to identify the bivariate model--fixing one of the β_m and one of the $\beta_{m,m+1}$ will suffice.

Rasch Markov Model Performance Evaluation

Discrete Bayes Validity analysis.

A brief explanation for the use of Bayes validities in comparing *competing models seems* necessary at this point. The use of Bayes estimate validities offers several advantages. First, Bayes estimates' performances reflect how the best possible estimation procedure for a given model will perform in terms of population validity (i.e. product-moment latent-trait, estimate correlation) as well as squared-error loss (Rao, 1965). Second, squared Bayes validities offer the useful interpretation as the maximum accountable variation reduction that is possible for any estimation scheme. Third, latent trait, maximum likelihood estimate correlations converge to Bayes validities the number of items increases. The overriding advantage, however, is that Bayes validities are much easier to obtain than other potentially better alternatives in studies such as this one, where models are at an early stage of development. For example, one superior approach to Bayes validity evaluation would be to validate an existing effective learning test, but such a test is not yet available. Another superior approach might be to compare the validities of estimates based on weaker assumptions than Bayes estimates, such as maximum likelihood estimates. However, such a comparison would require a prohibitively expensive Monte Carlo study, because existing maximum-likelihood algorithms for effective learning models are quite slow. What is needed instead is a preliminary analysis of learning effectiveness models in order to determine whether or not the expense and effort associated with future development and evaluation are worthwhile. The Bayes validity comparison approach has been chosen to fill that need.

Consider a (possibly vector-valued) statistic, S, having a finite number of possible values, s_1 through s_J, and depending on some N-variate latent

trait, $\underset{\sim}{\Theta}$, also taking on a finite number of possible values. (In practice, the finiteness limitation for Θ is minor insofar as continuous distributions may be reflected by discrete approximations.) If the latent trait probabilities,

$$h_{k_1,\ldots,k_N} = \Pr\{\Theta_1=\theta_{1k_1},\ldots,\Theta_N=\theta_{Nk_N}\},$$

are known along with the $\Pr\{S=s_j|\underset{\sim}{\Theta}=\underset{\sim}{\theta}\}$, then for $n=1,\ldots,N$, it is easy to compute the Bayes estimates,

$$b_{nj} = E[\Theta_n|S=s_j] = \frac{\sum_{k=1}^{K_n} \theta_{nk}\Pr\{S=s_j|\Theta_n=\theta_{nk}\}h_{nk}}{\sum_{k=1}^{K_n} \Pr\{S=s_j|\Theta_n=\theta_{nK}\}h_{nk}}, \tag{93}$$

where

$$h_{nk} = \sum_{\substack{k_1,\ldots,k_N \\ k_n=k}} h_{k_1,\ldots,k_n},$$

Given the Bayes estimates, it is also easy to directly compute the Bayes validities (Rao, 1965),

$$\mathrm{Corr}[B_n,\Theta_n] = \{\mathrm{Var}[B_n]/\mathrm{Var}[\Theta_n]\}^{1/2}, \tag{94}$$

and the expected proportions of accountable variation,

$$\frac{\mathrm{Var}[\Theta_n] - E\{\mathrm{Var}[\Theta|S]\}}{\mathrm{Var}[\Theta_n]} = \mathrm{Corr}[B_n,\Theta_n]^2.$$

In comparing the Bayes validities of competing models, two types of specification are needed, specifying each assumed model and specifying each true model. In effect, choosing the statistic, S, to be used in (93) corresponds to specifying the assumed model, because S will be sufficient for

Θ if and only if the assumed model corresponding to S holds. On the other hand, choosing the kernel in (93) along with the latent trait probability function, h, correspond to specifying the true model, because these govern the probabilities for S, whether or not S is the true sufficient statistic. The resulting basis for model comparisons is a table of Bayes validities, one for each assumed model-true model pair. Each entry gives the best validity that could be obtained for a given true model if the sufficient statistic associated with an assumed model were used.

Thus, in comparing the Bayes validities among the competing models for this article, basing validity computations on S = G, S = T, and S = (G,D) will yield Bayes validities associated with assuming additive Rasch, univariate Rasch Markov, and bivariate Rasch Markov models respectively. On the other hand, true models will be specified in terms of the appropriate probabilities for the possible g, t, and (g,t) values, if the true underlying models were additive Rasch, univariate Rasch Markov, and bivariate Rasch Markov. The resulting basis for model comparison is Table 2. Specific computational formulas for the Table 2 cases follov.

Starting with a description of the latent trait distributions for the Table 2 models, the univariate latent trait probabilities be discrete on 5 points, -2, -1, -0, 1, and 2, with probabilities based on a partition of the standard normal random variable, Z, yielding $Pr\{\Theta=-2\} = Pr\{Z<-1.5\} = 0.07$, $Pr\{\Theta=-1\} = Pr\{-1.5<Z<-0.5\} = 0.24$, and so on. For bivariate cases, individual latent trait values are also defined on -2, -1, 0, 1, and 2, with the same corresponding marginal probabilities as in the univariate case. The univariate Rasch Markov model is specified by formulating bivariate Rasch probabilities with $\Theta \equiv \gamma \equiv \delta$ on the 5 points; bivariate Rasch Markov probabilities with $\rho_{\gamma\delta} = -1$ are specified by setting $\gamma = -\delta$ on the 5 points;

and bivariate Rasch Markov probabilities with $\rho_{\gamma\delta} = 0$ are specified by setting $Pr\{\gamma,\delta\} = Pr\{\gamma\}Pr\{\delta\}$ on the 5 points.

Turning next to the conditional distributions corresponding to the true models in Table 2, if an additive Rasch model holds (first column in Table 2) the computational formulas for sufficient statistic probabilities are as follows. For $g = 0,\ldots,M$, $d = 0,\ldots,M-1$, and $\theta = -2,-1,\ldots,2$,

$$Pr\{g,d\,|\,\theta\} = \nu(\theta)c_{gd}\exp\{g\theta\},$$

where the c_{gd} are the joint contingencies among g and d (Appendix 1),

$$\nu(\theta) = \left[\sum_{g=0}^{M}\sum_{d=0}^{M-1} c_{gd}\exp\{g\theta\}\right]^{-1}$$

$$Pr\{g,d\} = \sum_{\theta=-2}^{2} Pr\{g,d\,|\,\theta\}h_\theta$$

$$Pr\{g\,|\,\theta\} = \sum_{d=0}^{M-1} Pr\{g,d\,|\,\theta\}$$

$$Pr\{g\} = \sum_{\theta=-1}^{2} Pr\{g\,|\,\theta h_\theta\}$$

$$Pr\{d\,|\,\theta\} = \sum_{g=0}^{M} Pr\{g,d\,|\,\theta\}$$

$$Pr\{d\} = \sum_{\theta=-2}^{2} Pr\{d\,|\,\theta\}h_j$$

$$Pr\{t\,|\,\theta\} = \sum_{g=0}^{t}\sum_{d=0}^{t-g} Pr\{g,d\,|\,\theta\}$$

and

$$Pr\{t\} = \sum_{\theta=-2}^{2} Pr\{t\,|\,\theta_k\}h_j \ .$$

Once the above values are computed, Bayes estimates are obtained from (93) and validities are obtained from (94), for each assumed model, that is for s = g, t, and (g,t).

If a bivariate Rasch model holds, the computational formulas for sufficient statistic probabilities are slightly different. For $g = 0,\ldots,M$, $d = 0,\ldots,M-1$, and $\gamma,\delta = -2,-1,\ldots,2$,

$$\Pr\{g,d\,|\,\gamma,\delta\} = \nu(\underset{\sim}{\theta})c_{gd}\exp\{g\gamma+d\delta\} ,$$

where

$$\nu(\underset{\sim}{\theta}) = \left[\sum_{g=0}^{M}\sum_{d=0}^{M-1} c_{gd}\exp\{g\gamma+d\delta\}\right]$$

$$\Pr\{g,d\} = \sum_{\gamma=-2}^{2}\sum_{\delta=-2}^{2} \Pr\{g,d\,|\,\gamma,\delta\}$$

$$\Pr\{g\,|\,\gamma\} = \frac{\sum_{\delta=-2}^{2} \Pr\{g,d\,|\,\gamma,\delta\}h_{\gamma\delta}}{\sum_{\delta=-2}^{2} h_{\gamma\delta}} ,$$

and so on. As in the univariate case, Bayes estimates and validites are obtained from (93) and (94), for each assumed model. For the bivariate case, however, three distinct (θ,δ) distributions are used, namely those that are determined by the above marginal probabilities and the stipulations that $\rho_{\gamma\delta} = 1$, 0, and -1.

The resulting Bayes validities that appear in Table 2 have been discussed earlier

Estimator accuracy analysis.

Simulations were performed in order to assess the performance of both maximum likelihood and ad hoc estimation procedures, based on both univariate and bivariate Rasch Markov models and sample sizes of both 5,000 and 50,000. Data for the simulations were generated in two stages: the first stage entailed randomly generating each individual's latent trait value(s), and the second stage entailed randomly generating a test pattern for each

individual's previously generated latent trait value(s). For univariate Rasch Markov simulations, θ values were generated from a discrete distribution on five points--the same univariate latent trait distribition that was used for table 2. For bivariate Rasch Markov simulations, the same discrete distribution for γ, δ was used as the discrete distribution in Table 2 corresponding to $\rho_{\gamma\delta} = 0$. For univariate or bivariate latent trait values generated in stage 1, test patterns were generated based on a 10-item test with $\beta_1 = \ldots = \beta_{10} = 1.0$ and $\beta_{12} = \ldots = \beta_{M-1,M} = -0.5$. Test patterns for univariate and bivariate simulations were generated in accordance with (9) and (69), respectively.

For each simulation, as test patterns were generated individual and item sufficient statistic values were compiled. For both item and individual parameters, all subsequent estimation was based exclusively on sufficient statistic values and frequencies, rather than estimating parameters separately for each individual. (Estimating only for distinct sufficient statistic values makes estimation computing time constant, no matter how many individuals are in a sample.)

In addition to using Newton-Rapheson iterations, steepest ascent iterations were used, because second derivatives/hessians for individual and item parameter estimation are not necessarily negative/negative definite away from likelihood maximizing values, given the models under study. In particular, at each estimation step, the value of each eigenvalue for the appropriate hessian was computed. If all eigenvalues were negative then Newton-Rapheson estimation was used; otherwise steepest ascent was used. Also, because calculations of the "symmetric functions," (44) through (56) and (82) through (90), can lead to inaccuracies for extreme parameter values, several preventive techniques were used. These included (a) steepest-ascent

line searches and bisection line searches in Nevton-Rapheson directions, based on small, preset maximum step sizes; and (b) the use of certain dual computations when calculating derivatives, in order to avoid overflow inaccuracies. For example in computing the gradient (29), if θ were large both the numerator and denominator would become large, leading to potential overflov in computing the τ_{ij} in (47) for large M, i, and j. This potential problem is avoided by focusing instead in such cases on the number of items that were answered incorrectly. Resulting dual computations involve converting elements of $\underset{\sim}{\lambda}$ to their reciprocal values, yielding progressively small values of the τ_{ij} as i and j increase. Given the resulting diminishing instead of increasing terms, when algorithms for evaluating the τ_{ij} find computed values to be sufficiently small computations may be halted for higher i, j values, with negligible loses in accuracy.

For both univariate and bivariate simulations, estimates were obtained in line with the maximum likelihood and ad hoc procedures that have been previously described. For individual parameter estimation the convergence requirement was that all item parameter gradients be less than 0.01 in magnitude. For item parameter estimation the convergence requirement was that all item parameter gradients be less in magnitude than 0.001 times the number of individuals (since item log-likelihoods are proportional to sample sizes). For all simulations, individuals having sufficient statistics at boundary values (e.g. t = 0 and t = 2M-1 for univariate Rasch Markov simulations) were not included in summary statistic computations. Finally, for each simulation parameter estimates were centered such that each average individual parameter estimate corresponding to a true value of 0 was also 0.

Table 3 contains summary statistics from each of the 8 sumulations (2 methods × 2 sample sizes × univariate vs. bivariate models that were

performed). Beginning with the univariate results, all simulations yielded product-moment correlations of about 0.81 between estimated and true θ values. These correlations compare favorably to the optimum Bayes validity of 0.91 (as obtained analytically in accordance with the previously described validity analysis), in light of individuals with extreme observed scores having been excluded. From the univariate individual estimate means in Table 3, it appears that estimated θ values are biased toward their mean of 0. This is due at least in part to extreme observed scores having been deleted. For example, more individuals having θ values of -2 tend to have $t = 0$, thus being deleted, than individuals having θ values of 0. This tends to make estimated θ means higher for the -2 group than they would have been, had a small estimated θ value been associated with t scores of 0.

Two kinds of bias are evident in the item parameters from the univariate Table 3 results. First, the additive item parameters tend to be separated from the cross-product item parameters by less than the corresponding true separation of 1.5, especially for results based on maximum likelihood. This bias is probably related to well-known similar problems asociated with maximum likelihood parameter estimation for the additive Rasch model (Andersen, 1980, pp. 240-245). The second type of bias is a shifting of all item parameter estimates either in the positive (for maximum likelihood results) or the negative (for ad hoc results) direction. This bias is at least partly caused by excluding extreme-valued individuals from the analyses. For example, when more individuals tend to have scores of $t = 2M-1$ than $t = 0$, maximum likelihood θ estimates will tend to be lower than they should be, and centering the results as in Table 3 will tend to make the $\underline{\beta}$ estimates higher than they should be.

Turning next to the bivariate Table 3 results, true-estimated correlations for these models are around 0.53 for γ and 0.38 for δ, somewhat lower than their maximum possible validity of 0.64 for γ and much lower than the maximum possible validity of 0.73 for δ. The probable main cause for this relatively poor bivariate estimation performance is that many more observations are lost due to boundary-valued sufficient statistics than in the univariate case. For example, if M = 10 and d = 2, then g values of 3 and 6 must be excluded because they are at extreme possible g values given that d = 2 and as a result they lead to diverging γ estimates. Many such constraints hold for g given d as well as for d given g, resulting in a large proportion of missing values if maximum likelihood is used and M is small. Boundary problems seem to dominate the bivariate item parameter estimate results in Table 3 as well, especially for the maximum likelihood method.

Related Models

Rasch Markov models are among the simplest members within the broad locally dependent/conjunctive item response model family (Jannarone, 1986). One way to describe the relative simplicity of Rasch Markov models is in terms of sufficient statistics for serial dependence (Kempf, 1977). For example, one interpretation from (18) is that given the Rasch Markov model (9), x_{s-1} is sufficient for describing the dependence of x_1 through x_{s-1} on $x_s (s=2,\ldots,M)$. Developments similar to (18) can be made for more elaborate conjunctive Rasch models, yielding correspondingly more elaborate sufficient statistics for serial dependence. For example, if the globally sufficient statistic for θ were

$$\sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} x_n x_{n+1} + \sum_{p=1}^{M-2} x_p x_{p+2},$$

then $x_{s-1} + x_{s-2}$ would be a sufficient statistic for the dependence of x_s on x_1 though x_{s-1}; if the globally sufficient statistic were

$$\sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} x_n x_{n+1} + \sum_{p=1}^{M-2} x_p x_{p+2} + \sum_{q=1}^{M-3} x_q x_{q+1} x_{q+2},$$

then the dependence sufficient statistic would be $x_{s-1} + x_{s-2} + x_{s-1}x_{s-1}$, and so on. Thus, among the broad class of serially dependent models, Markov models are perhaps the simplest, yielding a dependence sufficient statistic that is merely that single, most recent preceding event's value.

Continuing in terms of sufficient statistics for serial dependence, one Rasch conjunctive model is closely related to Kempf's (1977) dynamic test model, which in turn seems closely related to a locally dependent model proposed by Andrich (1985). It can be shown that the Rasch conjunctive kernel having

$$\sum_{m=1}^{M} x_m + \sum_{n=1}^{M-1} \sum_{p=n+1}^{M} x_n x_p \qquad (95)$$

as a sufficient statistic for Θ has

$$\sum_{t=1}^{s-1} x_t$$

as a sufficient statistic for the dependence of x_s on x_1 through x_{s-1}. The sufficient statistic (95) corresponds to a potentially interesting model, because as Kempf (1977) and others have pointed out there are many applications where <u>all</u> preceding events in a series could be utilized in a given trial. The corresponding Rasch conjunctive kernel does not appear to be as interesting as Kempf's, however, because the values of its sufficient statistic are related by a monotone one-to-one function onto the values of the additive Rasch Model's sufficient statistic given by (2). Consequently, (95) is equivalent in a sense to the additive Rasch model.

Along different lines, analogs to Rasch Markov models may also be formulated for the binomial and logistic conjunctive subfamilies. Markov counterparts to these models are the following special cases of equations (22) and (23) respectively in Jannarone (1986). The binomial extension is

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\} = \int_0^1 \varepsilon(\pi) \prod_{m=1}^{M} \pi^{x_m} \prod_{n=1}^{M-1} \pi^{x_n x_{n+1}} dU(\pi) , \qquad (96)$$

and

$$\Pr\{\underset{\sim}{X}=\underset{\sim}{x}\} = \int_{-\infty}^{\infty} \zeta(\Theta, \underset{\sim}{\alpha}, \underset{\sim}{\beta}) \exp\left\{ \sum_{m=1}^{M} \alpha_m(\Theta-\beta_m)x_m \right.$$

$$\left. + \sum_{n=1}^{M-1} \alpha_{n,n+1}(\Theta-\beta_{n,n+1})x_n x_{n+1} \right\} dV(\Theta). \qquad (97)$$

A Markov extension of the classical test model may also be formulated as the following special case of equation (27) in Jannarone (1986). For $\underset{\sim}{v}^T = (x_1,\ldots,x_M,x_1x_2,\ldots,x_{M-1,M})$, let

$$f(\underset{\sim}{x}|\theta) = \eta(\theta,\sigma_e^2)\exp\{-(\underset{\sim}{v}-\underset{\sim}{1}_{2^M}\theta)^T\underset{\sim}{V}_{\underset{\sim}{v}}^{-1}(\underset{\sim}{v}-\underset{\sim}{1}_{2^M}\theta)/2\} \quad , \tag{98}$$

where $\underset{\sim}{V}_{\underset{\sim}{v}} = \underset{\sim}{I}_{2^M}\sigma_e^2$ and

$$\eta(\theta,\delta_e^2) = \left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\{-(\underset{\sim}{t}-\underset{\sim}{1}_{2^M}\theta)^T\underset{\sim}{V}_{\underset{\sim}{v}}^{-1}(\underset{\sim}{t}-\underset{\sim}{1}_{2^M}\theta)/2\}du_1,\ldots du_M\right]^{-1} ,$$

with $\underset{\sim}{t}^T = (u_1,\ldots,u_M,u_1u_2,\ldots,u_{M-1}u_M)$.

Bivariate versions of (96) through (98) could also be constructed in a similar way to the bivariate Rasch Markov construction described earlier.

Discussion

Future work. Although the necessary models and estimation framework have been mapped out for effective learning tests in this article, the necessary work for making such tests a practical reality has just begun. At this point the algorithms for maximum likelihood estimation are prohibitively slow and possibly inaccurate, whereas only an ad hoc conditional maximum likelihood approach has been described. Necessary improvements to be worked out include the tailoring of known solutions for similar problems in the Rasch model (Andersen, 1982; de Gruijter, 1985; Formann, 1986; Gustaffson, 1980; Wainer, Morgan & Gustaffson, 1980) to such problems in Rasch Markov models. In particular, developing a conditional maximum likelihood method could result in less biased estimates, as could utilizing individual sufficient statistics having boundary values (perhaps by imposing benign Bayesian structures (Mislèvy, 1986) on latent trait distributions). Also, exploring simpler models and methods for effective learning models (Jannarone & Roberts, 1984) may be fruitful.

Other necessary work will include constructing effective learning tests and validating them empirically. Work has begun toward modifying and extending some existing effective learning-like items (McCormick & Miller, 1986) so as to fit a Rasch Markov model. Natural subsequent steps would include developing and evaluating Rasch Markov analogs for settings where conjunctive/locally dependent models seem promising, including discrete discrimination (Goldstein & Dillon, 1978; Young & Calvert, 1974), structural learning models (Rumelhart, 1986), and certain choice models (Laughlin & Jannarone, 1986).

Summary

A family of locally dependent models has been presented with potential applications in learning ability assessment and other areas. Bases for efficient estimation have been provided for two special cases--a univariate Rasch Markov model having a single individual parameter, and a bivariate Rasch Markov model having separable "aptitude" and "learning" components. Optimum validity analysis results have also been provided, showing some potential for improved test construction based on Rasch Markov models as well as potential redundancies, relative to traditional locally independent models. Finally, a small number of simulation results have been presented, pointing out several estimation problems that require future attention.

References

Andersen, E. B. (1980). Discrete statistical models with social science applications. Amsterdam: North Holland.

Andersen, E. B. (1982). Latent trait models and ability parameter estimation. Applied Psychological Measurement, 6, 445-461.

Andrich, D. (1985). A latent trait model for items with response dependencies: implications for test construction and analysis. In S. E. Embretson (Ed.), Test Design. Orlando, FL: Academic Press.

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., and Steinberg, E. (1985). Breakdowns in flexible uses of information: intelligence-related differences in transfer following equivalent learning performance. Intelligence, 9, 297-315.

de Gruijter, D. N. M. (1985). A note on the asymptotic variance-covariance matrix of item parameter estimates in the Rasch model. Psychometrika, 50, 247-249.

Embretson, S. E. (1985a). Multicomponent latent trait models for test design. In S. E. Embretson (Ed.), Test Design. Orlando, FL: Academic Press.

Embretson, S. E. (1985b, June). Component latent trait models: a remedy for local dependence? Paper presented at the annual Psychometric Society meeting, Nashville.

Fischer, G., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. Applied Psychological Measurement, 6, 397-416.

Formann, A. K. (1986). A note on the computation of the second-order derivatives of the elementary symmetric functions of the Rasch model.

Psychometrika, 51, 335-340.

Goldstein, M., & Dillion, W. R. (1978). Discrete Discriminant Analysis. New York: Wiley.

Groner, R. & Spada, H. (1977). Some Markovian models for structural learning. In M. Spada & W. F. Kempf (Eds.), Structural Models of Thinking and Learning. Bern: Hans Huber.

Gustafsson, J.-E. (1980). A solution to the conditional estimation problem for long tests in the Rasch model for dichotomous items. Educational and Psychological Measurement, 40, 377-385.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. Psychometrika, 51, ____ - ____ .

Jannarone, R. J. & Laughlin, J. E. (1986). Differential individual sensitivities to item characteristics. In preparation.

Jannarone, R. J. & Roberts, J. S. (1984). Reflecting interactions among personality items: Meehl's paradox revisited. Journal of Personality and Social Psychology, 47, 621-628.

Kempf, W. F. (1977). A dynamic test model and its use in the microevaluation of instructional material. In H. Spada & W. F. Kempf (Eds.), Structural Models of Thinking and Learning. Vienna: Hans Huber.

Laughlin, J. E. & Jannarone, R. J. (1986, March). A comprehensive measurement model for scanner-based choice data. ORSA/TIMS annual meetings, Dallas.

Lehmann, E. L. (1983). Theory of Point Estimation. New York: Wiley.

Lord, F. J. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. J. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

McCormick, C. & Miller, G. E. (1986). A comparison of mnemonic approaches to learning English vocabulary. In preparation.

Mislevy, R. J. (1986). Bayes model estimation in item response models. Psychometrika, 51, 177-196.

Rao, C. R. (1965). Linear Statistical Inference and its Applications. (2nd Ed.), New York: Wiley.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press.

Rumelhart, D. E. & McClelland, J. L. On learning the past tenses of English verbs. In McClelland, Rumelhart, & (Eds.) Parallel Distributed Processing: Explorations in the Microstructures of Cognition: Vol. 2, Psychological & Biological Models. Cambridge, MA: Bradford Books/MIT Press.

Spada, H. & McGaw, B. (1985). The assessment of learning effects with linear logistic test models. In S. E. Embretson (Ed.), Test Design. Orlando, FL: Academic Press.

Spray, J. A. & Ackerman, T. A. (1986, June). The effect of item response dependency on trait or ability dimensionality. Paper presented at the annual Psychometric Society Meetings. Toronto.

Sympson, J. B. (1977, July). A model for testing with multidimensional data. Paper presented at the University of Minnesota Computerized Adaptive Testing Conference.

Wainer, H. Morgan, A. & Gustafsson, J.-E. (1980). A review of estimation procedures for the Rasch model with an eye toward longish tests. Journal of Educational Statisitcs, 5, 35-64.

Young, T. Y. & Calvert, T. W. (1974). Classification, Estimation and Pattern Recognition. New York: American Elsevier.

Appendix 1. Rasch Markov sufficient statistic contingencies.

The number of distinct test patterns, c_{gd}, leading to bivariate Markov joint sufficient statistic values (g,d) (g=1,2,...,M; d=0,1,...,M-1) can be described as follows. One such test pattern takes the form,

(1,0,1,0,...,1,0,1,1,...,1,0,0,...,0),

where k = g-1-d zeros are flanked by the first k+1 ones, followed by the remaining d ones, followed by the remaining M-g-k zeros. From the above basic pattern there are two variants that will preserve a given (g,d) value. First, within the first M-g basic pattern elements the individual zeros may be positioned between any two adjacent ones. There are

$$\binom{t-1}{k}$$

such configurations including the basic pattern. Second, for each such configuration it is possible to position any of the remaining M-g zeros either to the left of the first one, to the right of the last one, or between any of the separated ones, without changing (g,d). The number of such positions, including the basic pattern, is,

$$\binom{M-g-k+1}{k+1} .$$

It follows that the contingencies of distinct patterns for given (g,d) values are,

$$c_{gd} = \binom{g-1}{k}\binom{M-g-k+1}{k+1},\ g=1,\ldots,M,\ k=0,\ldots,\min(g-1,M-g),$$

$$= 1,\ g=d=0,$$

$$= 0 \text{ elsewhere.}$$

Table 1.

Contingencies among Test score patterns yielding Distinct Joint and Marginal g,d values (M=15)*

		d														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	marginal g frequencies
	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15
	2	91	14	0	0	0	0	0	0	0	0	0	0	0	0	0	105
	3	286	156	13	0	0	0	0	0	0	0	0	0	0	0	0	455
	4	495	660	198	12	0	0	0	0	0	0	0	0	0	0	0	1365
	5	462	1320	990	220	11	0	0	0	0	0	0	0	0	0	0	3003
	6	210	1260	2100	1200	225	10	0	0	0	0	0	0	0	0	0	5005
	7	36	504	1890	2520	1260	216	9	0	0	0	0	0	0	0	0	6435
g	8	1	56	588	1960	2450	1176	196	8	0	0	0	0	0	0	0	6435
	9	0	0	28	392	1470	1960	980	168	7	0	0	0	0	0	0	5005
	10	0	0	0	0	126	756	1260	720	135	6	0	0	0	0	0	3003
	11	0	0	0	0	0	0	210	600	450	100	5	0	0	0	0	1365
	12	0	0	0	0	0	0	0	0	165	220	66	4	0	0	0	455
	13	0	0	0	0	0	0	0	0	0	0	66	36	3	0	0	105
	14	0	0	0	0	0	0	0	0	0	0	0	0	13	2	0	15
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
marginal d frequencies		1597	3970	5807	6304	5542	4118	2655	1496	757	326	137	40	16	2	1	Total 32,768

*Each entry is the number of distinct test patterns from a 15-item test that could yield the indicated joint and marginal and d values.

Table 2.

Optimal Performance Characteristics of Individual Parameter Estimators*

Assumed Model (sufficient statistic)	Number of Items	Actual Probability Structure† (Univariate) Additive Rasch	Univariate Rasch Markov	Markov, $\rho_{\gamma\delta}=0$	Markov, $\rho_{\gamma\delta}=-1.0$
Additive	15	87	93	64,69	32
Rasch	27	93	96	71,67	32
(g)	39	95	97	73,68	32
Univariate	15	87	93	66,67	30
Rasch	27	92	96	69,70	32
Markov (t)	39	95	97	70,71	35
Bivariate	15	87	93	73,74	71
Rasch	27	93	96	78,80	80
Markov (g,d)	39	95	97	82,83	85

* Entries are product-moment correlations (× 100) between actual latent traits and their Bayes estimates based on assumed sufficient statistics.

† All actual probability structures are based on Rasch Markov models with zero-valued item parameters and discrete latent traits. For univariate structures, θ values occur at -2, -1, 0, 1, and 2, with (quasinormal) probabilities, 0.07, 0.24, 0.38, 0.24, and 0.07, respectively. For bivariate structures, (γ, δ) values occur at (-2,-2), (-2,1),...,(2,2), having the same marginal probabilities as univariate θ values.

Table 3.

Accuracy Analysis Results for Two Estimation Procedures*

Procedure	Sample Size	Model	Individual Parameter-Estimate Correlations	Individual Parameter Estimate Means, True Parameter Values −2	−1	0	1	2	Item Parameter Estimate Means and Standard Deviations, Additive Parameters (True Values all 1)	Cross-Product Parameters (True Values all −0.5)
Maximum Likelihood										
	5,000	Univariate	0.81	−0.87	−0.69	0	1.00	1.18	0.90, 0.19	−0.07, 0.31
	5,000	Bivariate (γ)		−1.05	−0.66	0	0.60	1.15	−1.51, 0.97	
		Bivariate (δ)		−0.99	−0.47	0	0.38	0.60		−0.25, 0.95
	50,000	Univariate	0.81	−0.90	−0.70	0	0.98	1.15	0.87, 0.18	−0.08, 0.26
	50,000	Bivariate (γ)	0.53	−1.06	−0.68	0	0.61	0.98	−1.34, 1.10	
		Bivariate (δ)	0.39	−0.86	−0.50	0	0.40	0.63		−0.33, 0.73
Ad Hoc										
	5,000	Univariate	0.81	−0.95	−0.69	0	0.96	1.16	1.04, 0.18	−0.91, 0.38
	5,000	Bivariate (γ)		−1.14	−0.66	0	0.60	0.70	0.97, 0.30	
		Bivariate (δ)		−0.83	−0.46	0	0.39	0.69		−1.53, 0.24
	50,000	Univariate	0.81	−0.92	−0.71	0	0.97	1.12	1.10, 0.08	−0.84, 0.11
	50,000	Bivariate (γ)	0.53	−1.09	−0.67	0	0.39	0.69	0.99, 0.12	
		Bivariate (δ)	0.39	−0.88	−0.50	0	0.42	0.63		−1.55, 0.21

* Based on data generated from one univariate and one bivariate Rasch Markov model, both with 10 items having additive item parameters set to 1 and cross-product item parameters set to −0.5. Univariate and bivariate discrete distributions that were used to generate latent traits correspond respectively to the univariate and bivariate ($\rho_{\gamma\delta}$=0.) distributions in Table 2.

Figure 1. A Single-trait General and Learning Aptitude Model.

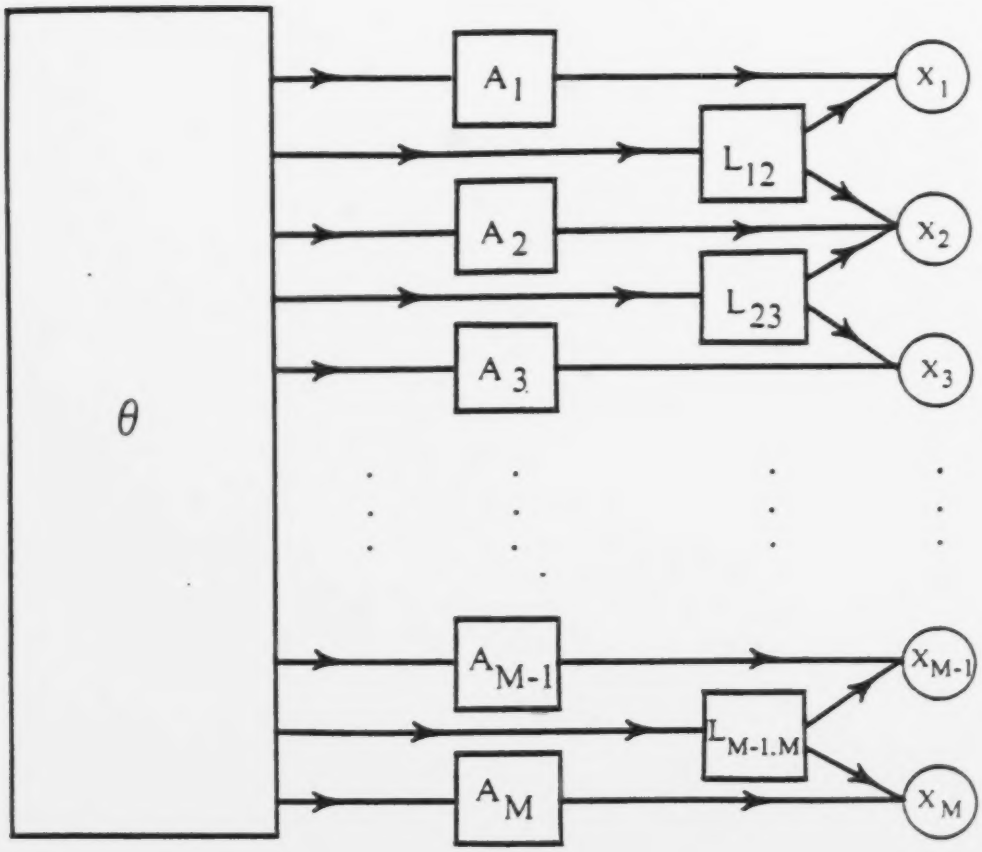


Figure 2. An Aptitude Model Having Distinct General and Learning Traits.

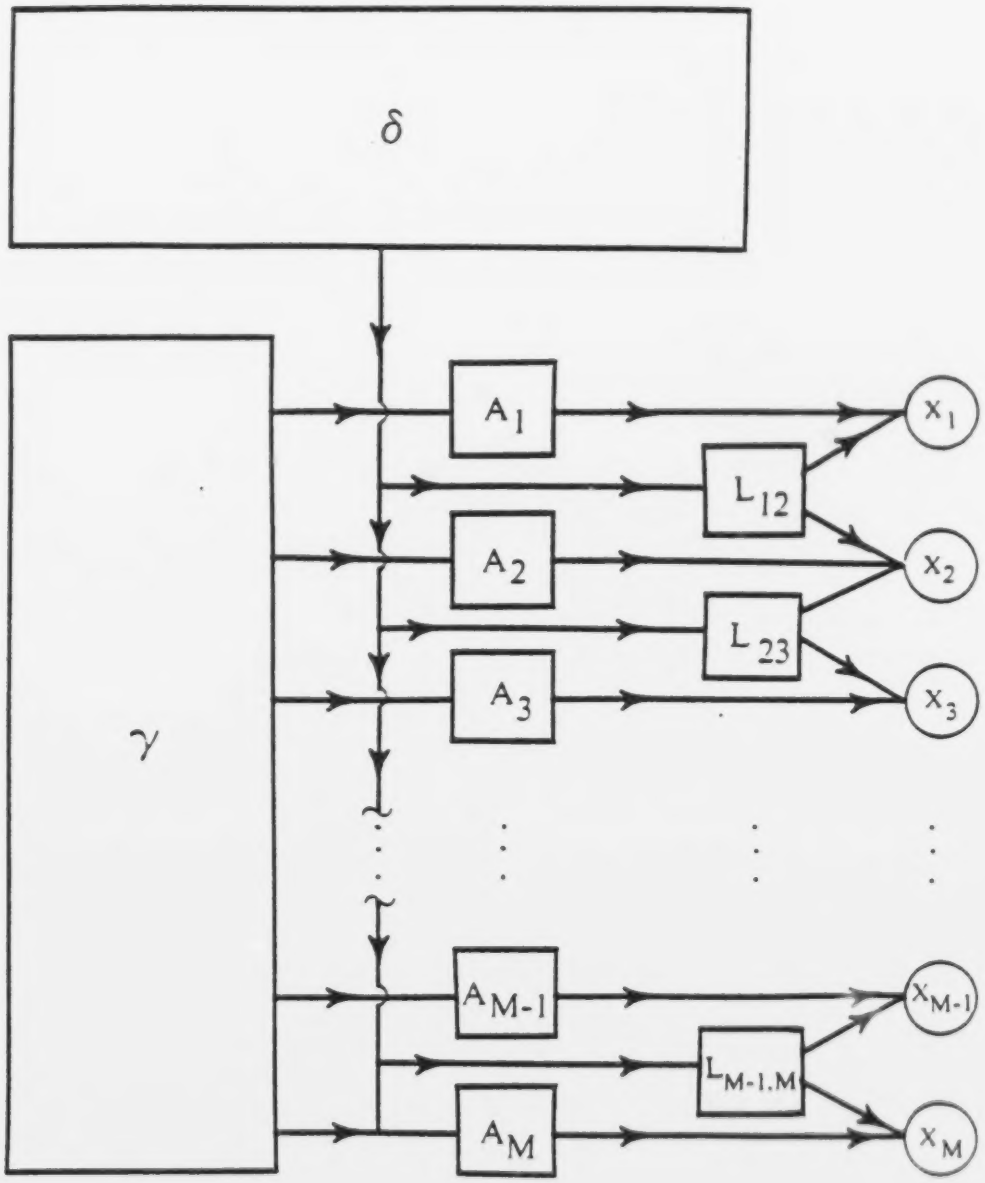
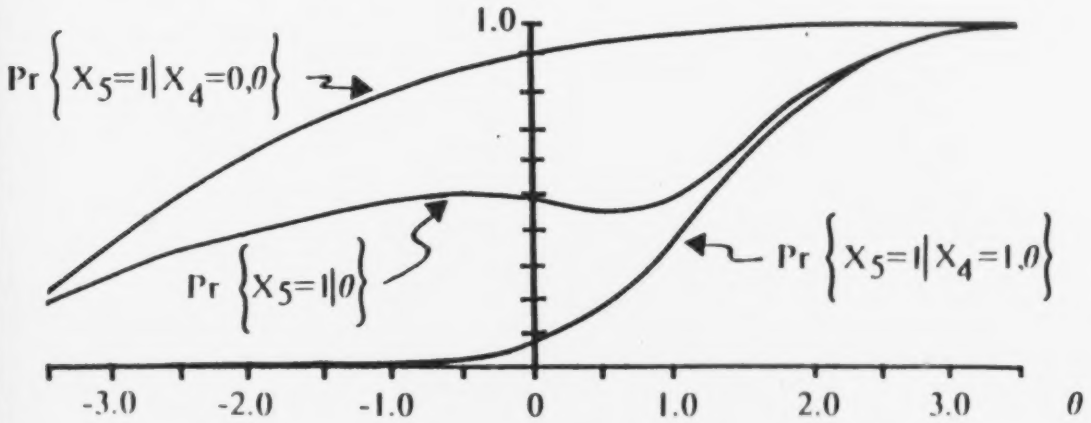


Figure 3. Response Probabilities For One Item, Given a Particular Univariate Rasch Markov Model.



* Based on item 5 from a 5-item test satisfying a Univariate Rasch Markov Model with $\beta_1=\beta_2=\beta_3=\beta_{12}=\beta_{23}=\beta_{34}=0$, $\beta_4=\beta_5=-2.5$, and $\beta_{45}=5.0$.

Personnel Analysis Division,
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Phipps Arabie
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramt Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

Dr. Bruce Bloxom
Administrative Sciences
Code 54B1
Navy Postgraduate School
Monterey, CA 93943-5100

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

Dr. Robert Brennan
American College Testing
Programs
P. . Box 168
Iowa City, IA 52243

Dr. Patricia A. Butler
OERI
555 New Jersey Ave., NW
Washington, DC 20208

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
OP 01B7
Washington, DC 20370

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Qincy Street
Arlington, VA 22217-5000

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Ball Foundation
800 Roosevelt Road
Building C, Suite 206
Glen Ellyn, IL 60137

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 25
San Antonio, TX 78228

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Carl H. Frederiksen
McGill University
3700 McTavish Street
Montreal, Quebec H3A 1Y2
CANADA

Dr. Robert D. Gibbons
University of Illinois-Chicago
P.O. Box 6998
Chicago, IL 69680

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Douglas A. Jones
P.O. Box 6640
Lawrenceville
NJ 08648

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
The NETHERLANDS

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Mr. Robert McKinley
University of Toledo
Department of Educational Psychology
Toledo, OH 43606

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

Dr. Melvin R. Novick
356 Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Director Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

Office of Naval Research,
Code 1142PT
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Special Assistant for Marine
Corps Matters,
ONR Code 00MC
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. Roger Pennell
Air Force Human Resources
Laboratory
Lowry AFB, CO 80230

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Dr. Robert Sasmor
HQDA DAMA-ARL
Pentagon, Room 3E516
Washngton, DC 20310-0631
USA

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152-6800

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940